

University of Groningen

A computational cognitive modeling approach to the development of second-order theory of mind

Arslan, Burcu

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Arslan, B. (2017). *A computational cognitive modeling approach to the development of second-order theory of mind*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

**A computational
cognitive modeling approach
to the development of
second-order theory of mind**

BURCU ARSLAN



A computational cognitive modeling approach to the development of second-order theory of mind

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 21 April 2017 at 11.00 hours

by

Burcu Arslan

born on 22 May 1980
in Ankara, Turkey

© Burcu Arslan, Groningen, the Netherlands, 2017.

Cover & layout design:  Lovebird design.
www.lovebird-design.com

Printing: Eikon+

ISBN printed version: 978-90-367-9693-4

ISBN digital version: 978-90-367-9692-7

Supervisors

Prof. L.C. Verbrugge

Prof. N.A. Taatgen

Assessment Committee

Prof. P. Hendriks

Prof. M.A.J. Raaijmakers

Prof. L. Schooler

To my grandparents Sevinç, Şeyh Ömer, Mehmet and Nazlı

Table of contents

Chapter 1:	Introduction	9
Chapter 2:	Five-year-olds' Systematic Errors in Second-order False Belief Tasks Are Due to First-order Theory of Mind Strategy Selection: A Computational Modeling Study	29
Chapter 3:	Five-year-old Children's Development of Second-order False Belief Reasoning Can Be Accelerated: An Evaluation of Different Feedback Methods	61
Chapter 4:	Syntactic Recursion Facilitates and Working Memory Predicts Recursive Theory of Mind	93
Chapter 5:	The Role of Simple and Complex Working Memory Strategies in the Development of First-order False Belief Reasoning: A Computational Model of Transfer of Skills	123
Chapter 6:	Cognitive Control Explains the Mutual Transfer Between Dimensional Change Card Sorting and First-order False Belief Understanding: A Computational Modeling Study on Transfer of Skills	137
Chapter 7:	Putting the Pieces Together	161
	Samenvatting	175
	References	189
	S1 Materials	202
	S2 Materials	204
	Publication List	216
	Acknowledgments	219

Chapter 1: Introduction

*In which we give an overview of this dissertation
and discuss the underlying theories and applied
methodologies.*



What should I write into this Introduction in order to attract your attention to read the whole dissertation or at least until the end of the Introduction? To achieve this goal, first of all, I should take your perspective by thinking about your background, and then I should find a way to provide enough information to make you understand the rest of this dissertation without putting you to sleep. In other words, I should use *theory of mind* reasoning (Premack & Woodruff, 1978) by appreciating that you might have different knowledge, beliefs and desires than my own.

Maybe you *think that I believe that I found a way* of getting your attention but you *think that I am wrong*. When I reason that “you *think that I am wrong*”, I use first-order theory of mind reasoning by attributing a mental state to you. Furthermore, I use second-order theory of mind reasoning when I reason that “you *think that I believe that I found a way*” by attributing a mental state to you who attributes a mental state to myself. Second-order mental state attribution is important in many different social situations, such as idiom understanding (Caillies & Le Sourn-Bissaoui, 2013), maintaining a strategic lie (Hsu & Cheung, 2013), and irony understanding (Filippova & Astington, 2008). For example, although John says “You sure are a great researcher”, John doesn’t really want Stefan to believe that he is a great researcher.

This dissertation is part of a project called “Cognitive systems in interaction: Logical and computational models of higher-order social cognition”, awarded to my first supervisor Rineke Verbrugge. The general motivation of the project is to provide a better understanding of higher-order theory of mind for the benefits of cognitive scientists, logicians and computer scientists. In the near future, humans will work together with artificial agents in daily life. Investigating the underlying mechanisms of the limitations in humans higher-order theory of mind reasoning will allow us to build more effective systems for better communication, collaboration and negotiation between these artificial agents and humans. To this end, I focused on children’s development of second-order theory of mind in terms of learning in decision making, transfer of skills, cognitive control, working memory and language.

In addition to contributing to developmental psychology and cognitive science in general, studying children’s development of theory of mind can contribute to one of the emerging research domains, which is child-robot interaction. Child-robot interaction aims to enhance children’s healthcare and education with interactive robots. Considering that children do not perceive robots just as running programs and that they attribute characteristics of living things to the robots (Belpaeme et al., 2013), it is important that the robots that interact with children “know” the limits of children’s theory of mind depending on their ages together with the underlying mechanisms of these limitations. Moreover, to

enhance the robots' social and cognitive abilities, it is important that they "know" the effective ways to improve children's theory of mind abilities. This dissertation aims to contribute to all above-mentioned fields by providing new insights about children's development of second-order theory of mind.

Before focusing on the specifics of children's development, it might be useful to have a general view about theory of mind. For this purpose, in the light of the literature, I will first try to provide short answers to a couple of general questions about theory of mind.

1.1. What is the limit of recursion in adults' theory of mind reasoning?

Adults use theory of mind recursively up to the fourth level (e.g., "You want me to believe that my mother thinks that my aunt believes that my cousin is getting married") to answer questions related to vignettes (Kinderman, Dunbar, & Bentall, 1998).

Kinderman et al. tested adults with five different vignettes, which involved complex social situations. The participants were expected to choose one of two possible answers for each question. While the proportion of incorrect answers was around 20% up to the fourth-level of theory of mind question, proportion of incorrect answers increased dramatically to 60% at the fifth level. Importantly, participants' lower performance at fifth-order theory of mind reasoning was not due to forgetting the story facts.

In Chapter 4, we invoke the *serial processing bottleneck hypothesis* (Verbrugge, 2009) in order to provide a possible explanation to the question why processing embedded mental states requires more complex working memory strategies than simply remembering facts.

1.2. Can adults apply higher-order theory of mind flawlessly in different contexts as well?

Previous studies have shown that adults do not apply second-order theory of mind flawlessly in strategic games (Flobbe, Verbrugge, Hendriks, & Krämer, 2008; Hedden & Zhang, 2002; Meijering, van Maanen, van Rijn, & Verbrugge, 2010; Meijering, Taatgen, van Rijn, & Verbrugge, 2014).

In order to investigate the effect of context on the performance of theory of mind, Flobbe et al. (2008) constructed a two-player sequential game in which a human player and the computer opponent were expected to jointly drive a car

by taking turns. Figure 1.1 depicts a screenshot of a second-order theory of mind phase of the game. There were three decision points represented by road junctions. The end points of the game were represented by the dead ends. The number of blue (dark gray) and yellow (light gray) marbles at each dead end represented each player's reward. The participants were instructed to maximize their own reward (i.e., attain the highest possible number of blue marbles) and were told that the computer opponent would try to do the same (i.e., attain the highest possible number of yellow marbles). At the first road junction, the human player had to make a decision (i.e., to go to the dead end and finish the game or to go straight to continue). If the human player chose to go straight, the computer opponent made a decision. If the computer decided to continue, the human player was expected to finish the game by choosing to go right or to go straight in order to attain the highest possible reward. Therefore, in order to get the highest possible reward, the human player was expected to apply second-order theory of mind (e.g., "the computer 'thinks' that I would go straight to get seven marbles, therefore, the computer would go to the dead end at the second road junction. So, I should go

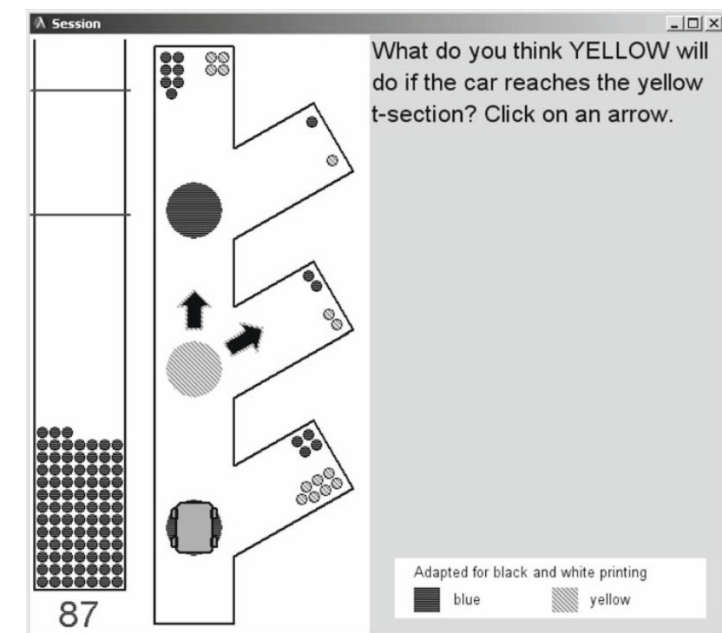


Figure 1.1. A screenshot of a second-order theory of mind phase of Flobbe et al.'s (2008) strategic game. Reprinted from "Children's application of theory of mind in reasoning and language," by Flobbe et al., 2008, *Journal of Logic, Language and Information*, 17 (4), p. 432. Copyright 2008 by the American Psychological Association.

to the dead end at the first road junction to attain the highest possible four marbles). The results showed that adults' performance was not perfect and 75% of the adults used the correct second-order theory of mind reasoning.

Moreover, it has been shown that adults have difficulties applying even first-order theory of mind in online communication games (Dumontheil, Apperly, & Blakemore, 2009; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003).

These findings together suggest that the context does matter and that use of theory of mind in a dynamic game is not an automatic and flawless process¹.

1.3. In which environments does use of higher-order theory of mind have an advantage?

Agent-based simulation research has provided elaborate answers for the question in which environments use of higher-order theory of mind might have an advantage. De Weerd, Verbrugge and Verheij (2013) have studied the function of higher-order theory of mind in competitive games. To this end, they compared computational agents' behavior in different game settings when the agents, who have different levels of theory of mind, play the games against other agents. Their results showed that using first-order and second-order theory of mind in a competitive game setting has an advantage over a competitor agent that does not take the perspective of the opponents and acts just based on the opponents' observed behavior.

Moreover, although agents were able to cooperate without using theory of mind, using first-order and second-order theory of mind in a cooperative setting allowed the agents come up with an agreement faster (de Weerd, Verbrugge, & Verheij, 2015). Finally, in all competitive, cooperative and mixed-motive situations, use of second-order theory of mind has an advantage over use of first-order theory of mind (de Weerd et al., 2013; 2015; 2017).

¹ However, for evidence of automatic visual perspective taking, see Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010.

1.4. Development of theory of mind

More than three decades of research have shown that theory of mind reasoning develops with age (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). It has been argued that being able to attribute a false belief to someone else provides evidence that a person has a theory of mind (Dennett, 1978). Since then, the false belief task has become a litmus test for theory of mind reasoning (Wellman et al., 2001). In a standard explicit false belief task, children are required to report a decision about another person's mental state while they know the real situation, which happens to be different from the other person's false belief.

In the rest of this section, I briefly review the development of theory of mind starting from infants' implicit false belief reasoning to young children's explicit first-order false belief reasoning, and finally, to the main focus of this dissertation, which is older children's development of second-order false belief reasoning².

1.4.1. Implicit false belief reasoning

A number of studies have found that infants can pass implicit first-order false belief tasks (Baillargeon, Scott, & He, 2010; Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005). Different from explicit false belief tasks, infants are tested with nonverbal false belief tasks in which infants' eye movements and looking times are measured.

In the implicit false belief task of Onishi and Baillargeon (2005), infants are first habituated by watching an adult who is putting a toy into a green box (Figure 1.2a, Trial 1) and reaching for the object in the green box (Figure 1.2a, Trials 2 and 3). Subsequently, infants are presented with the *belief-induction* trial where the toy is either moved from the green box to a yellow box while the adult is absent (Figure 1.2b, False-belief-green condition) or moved to the yellow box in the adult's presence but then moved back to the green box after the adult leaves (Figure 1.2b, False-belief-yellow condition). Finally, in the test trial, the adult returns and searches for the toy either in the yellow box (Figure 1.2c, Yellow-box event) or in the green box (Figure 1.2c, Green-box event). In the false-belief-green condition, the infants who saw the yellow-box event looked reliably longer than

² Note that children's development of theory of mind is not restricted to false belief reasoning and continues to develop after they pass second-order false belief tasks. A group of studies tested children with more naturalistic and advanced theory of mind tasks requiring application of first-order and second-order theory of mind (Devine & Hughes, 2013; Happé, 1994). They found that children start to pass these advanced tasks not before the age of eight. Moreover, it has been shown that it takes children a couple of more years to apply second-order theory of mind reasoning in strategic games after they pass second-order false belief tasks (Flobbe et al., 2008; Raijmakers, Mandell, van Es, & Counihan, 2014).

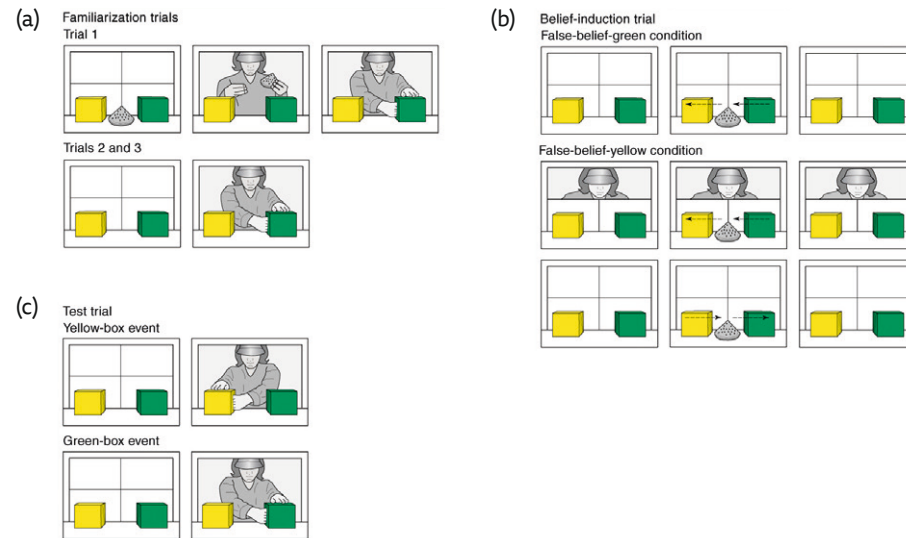


Figure 1.2. An implicit false belief task. a) Familiarization trials, b) belief-induction trial, c) test trial. Reprinted from “False-belief understanding in infants” by Baillargeon, Scott & He, 2010, *Trends in Cognitive Science*, 14(3), p. 111. Copyright 2005 by the American Psychological Association.

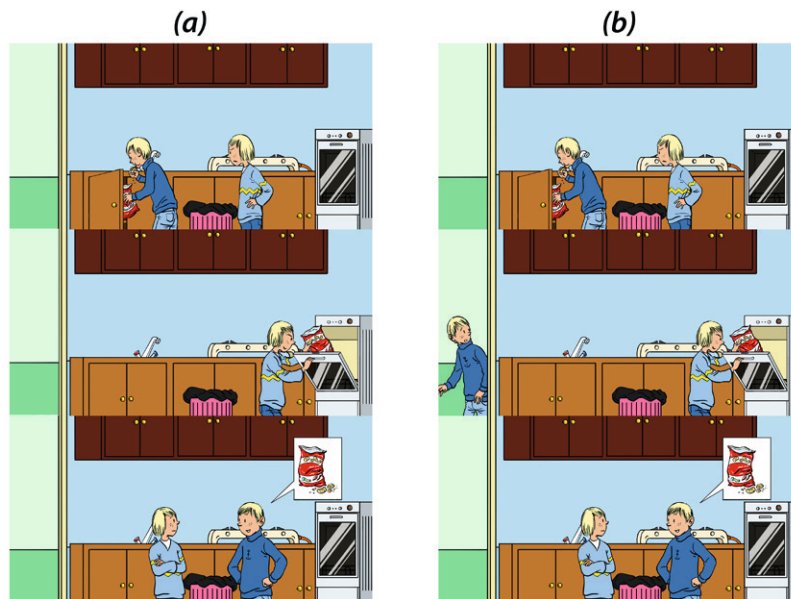


Figure 1.3. (a) An example of a first-order false belief task, (b) an example of a second-order false belief task (Illustration ©Avik Kumar Maitra).

those who saw the green-box event and, in the false-belief-yellow condition, the infants who saw the green-box event looked reliably longer than those who saw the yellow-box event.

The researchers concluded that infants do expect an agent to search for an object based on the agent's belief about the object's location, instead of the real location of the object. However, a group of researchers have argued that passing implicit false belief tasks does not necessarily require to have an implicit understanding of false beliefs and that behavioral reading is enough to pass the implicit false belief tasks (Heyes, 2014; Perner & Ruffman, 2005).

1.4.2. Explicit first-order false belief reasoning

To test explicit first-order false belief reasoning, children are presented with a story and expected to predict or explain the protagonist's action. Figure 1.3a depicts an example of an explicit first-order false belief task. Maxi eats some of his chips and puts the remainder into the cupboard. Once Maxi leaves the kitchen, Sally takes the chips from the cupboard and hides the bag of chips in the oven. After a while, Maxi comes back to the kitchen and says: “I want to eat my chips”. At this point, children are expected to answer the first-order false belief question, “Where will Maxi (first) look for his chips?”, often together with the justification question “Why does he look there?”.

While children around the age of four can correctly predict Maxi's false belief about the location of the chips by saying “in the cupboard”, most children around the age of three make systematic errors and give the answer “in the oven” without taking into account Maxi's ignorance about the real location of the chips (Wellman et al., 2001; but see Setoh, Scott, & Baillargeon, 2016 for evidence that two-and-a-half-year-olds can pass explicit first-order false belief tasks when the processing demands are reduced).

Why can most 3-year-olds not pass explicit first-order false belief tasks although infants show an implicit understanding of false belief reasoning? It has been proposed that human theory of mind reasoning involves two systems, which are independent from each other (Apperly and Butterfill, 2009; Low, Apperly, Butterfill, & Rakoczy, 2016). System 1 is fast, automatic and inflexible. On the other hand, System 2 is slow, effortful and flexible. While System 1 allows infants to pass implicit first-order false belief tasks, System 2 can only be tested explicitly and develops with age (see Carruthers, 2016a, 2016b for criticisms related to two-system accounts). Similar to this explanation, Heyes and Frith (2014) proposed that humans are born with a ‘start-up kit’. This ‘start-up kit’ involves genetic neurocognitive mechanisms that allow us to have accurate expectations

about behavior of other agents and allows infants to pass implicit theory of mind tasks. In contrast, they propose that explicit theory of mind is a culturally inherited skill which is learned by verbal instruction.

One of the proposed theories in order to explain 3- to 4-year-olds' transition from failure to success in the explicit first-order false belief tasks is the *conceptual change* theory (Gopnik & Wellman, 1994; Wellman et al., 2001)³. An explanation of the *conceptual change* theory is that children first construct naïve theories to predict others' behavior based on observable instances. However, these naïve theories do not always lead children to correct predictions, especially when others have false beliefs. Based on these incorrect predictions, just like scientists do, children accumulate evidence and revise their naïve theories to more elaborate theories that can also predict others' behavior based on unobservable instances (Goodman et al., 2006; Gopnik & Wellman, 1994). There are three major lines of research that support the view that 3-year-old children do need a *conceptual change* to pass first-order false belief tasks. The first one is related to the findings that children who have more siblings pass first-order false belief tasks earlier (Ruffman, Perner, Natio, Parkin, & Clements, 1998). The second one is related to the training studies showing that when children are trained on understanding the concept of belief, their performance on the first-order false belief tasks improves (Slaughter & Gopnik, 1996). Finally, it has been shown that it is not possible to accelerate children's first-order false belief reasoning without explicit feedback with further explanations (Clements, Rustin, McCallum, 2000). These studies together have been used as evidence for children's *conceptual change* because children do need to conceptually understand that people might have different mental states which can be different from their own mental states.

An important factor in children's development of explicit first-order theory of mind is language. Because explicit false belief tasks are verbal, it is not surprising that language is related to children's transitions from failure to success. However, the important question is in what ways language matters and which components of language are crucial in children's development of explicit first-order theory of mind (see Astington & Baird, 2005 for further details).

A group of researchers have shown that the *semantics* component of language is crucial for explicit theory of mind because the mental state verbs such as 'think', 'want', and 'believe' refer to unobservable states of the mind (Peterson & Siegel, 2000; Ruffman, Slade, & Crowe, 2002). Alternatively, another group of researchers have proposed that *syntax* is the crucial component of language, because complement clauses that involve mental state verbs, such as in "Maxi *thinks that* the bag of chips is in the cupboard", allow us to represent states that contrast

³ For simulation theory and hybrid theories of theory of mind, see Harris (2002), Goldman (2006), and Leslie, Friedman & German (2004).

with reality or with other people's mental states in terms of truth-value (de Villiers & Pyers, 2002; de Villiers, 2005; de Villiers, 2007). For example, in Figure 1.3a, while the statement "The bag of chips is in the cupboard" is false, the whole sentence "Maxi *thinks that* the bag of chips is in the cupboard" is true. A longitudinal study on children's false belief reasoning and language has shown that both semantics and syntax contribute to children's first-order false belief reasoning (Slade & Ruffman, 2005)⁴.

In contrast to the *conceptual change* theory, it has been claimed that children's failure in explicit first-order false belief tasks is mostly due to the *complexity* of the tasks in terms of executive functions (Leslie, 1994; Scholl & Leslie, 1999; Baillargeon et al., 2010). Executive function is an umbrella term that covers a set of cognitive processes, such as working memory, inhibition and cognitive control. The *complexity* camp proposes that children's failure in explicit false belief tasks is due to children's inability to keep track of events and to infer and hold in mind the contrary beliefs of other agents. Their problems are related to working memory, as well as inhibiting one's own perspective, and to cognitive control (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Claxton, & Moses, 2014; Davis & Pratt, 1995; Gordon & Olson, 1998; Keenan, Olson, & Marini, 1998).

Moreover, although it is not possible to rule out that children go through a *conceptual change*, the findings that individual differences in children's executive functioning predict improvements in first-order theory of mind tasks signal that the complexity of the explicit first-order false belief tasks contributes to children's transitions from failure to success (Benson, Sabbagh, Carlson, & Zelazo, 2013; Sabbagh, Hopkins, Benson, & Flanagan, 2010).

1.4.3. Second-order false belief reasoning

In order to investigate children's further development of false belief reasoning after the age of four, children are tested with second-order false belief tasks (Braüner, Blackburn, & Polyanska, 2016; Hollebrandse, van Hout, & Hendriks, 2014; Perner & Wimmer, 1985; Sullivan et al., 1994).

Figure 1.3b depicts an example of a second-order false belief task. Different from the first-order false belief task, while Sally takes the chips from the cupboard and hides it in the oven, Maxi passes by the kitchen door and sees Sally hiding the chips in the oven. However, Sally does not see Maxi. After Maxi comes back to the kitchen and says: "I want to eat my chips", children are expected to answer the second-order false belief question, "Where does Sally think that Maxi

⁴ See Westra & Carruthers (2017) for a pragmatic account.

will look for his chips?”, together with the justification question “Why does she think that?”. Most children around the age of five cannot attribute a false belief to Sally who is attributing a belief to Maxi, while most 6- to 7-year-olds can (Perner & Wimmer, 1985; Sullivan et al., 1994; Miller 2009; 2012).

Why does it take children another one or two years to pass second-order false belief tasks once they already know that mental states of others might be different from one’s own? In line with the first-order theory of mind literature, two types of explanations have been proposed for children’s developmental lag between first-order and second-order false belief reasoning, namely *conceptual change* and *complexity* explanations (Miller 2009; 2012).

The conceptual change explanation suggests that children need to realize that mental states such as beliefs can have other beliefs and not just events in the world as their content and can be used recursively (e.g., “Sally thinks that [Maxi believes that [the chips are in the cupboard]]”). This view implies that children need to hear many examples of recursive mental state talk in order to realize that mental states can be used recursively (Miller 2009; 2012).

On the other hand, the complexity explanation suggests that the higher complexity of second-order false belief reasoning adds further demands on executive functions, as does the linguistic complexity of the stories and the questions, in comparison to first-order false belief tasks.

As you may have already noticed when reading the previous subsection on children’s development of first-order false belief reasoning, there is an extensive research on that topic. On the other hand, the number of studies that focus on children’s development of second-order false belief reasoning is much more scant (see Miller, 2009; 2012 for a detailed review). Because some of these studies used composite scores of first-order and second-order false belief reasoning, we do not have an extensive knowledge on children’s developmental transitions from first-order to second-order false belief reasoning.

The Chapters 2,3 and 4 of this dissertation cover most of the studies on children’s development of second-order false belief reasoning. For this reason, I do not give detailed explanations about those studies here (see Chapter 2 for a literature review about the previous computational cognitive models of first-order and second-order false belief reasoning; see Chapter 3 for training studies on second-order theory of mind; and see Chapter 4 for studies related to language and executive functions in relation to second-order false belief reasoning). However, in the following section, I present the research questions of this dissertation, which have not been investigated before. Subsequently, I introduce the methodology that we used in order to find answers to these questions. Finally, I give an overview of the rest of the chapters of this dissertation.

1.5. Research questions

Why can children not pass second-order false belief tasks once they are able to pass first-order false belief tasks? The main goal of this dissertation is to provide a possible answer for this question by investigating the components of children’s development of second-order theory of mind. For this purpose, for the first time in the literature, we investigated the following research questions:

1. Once 5-year-old children already have zero-order and first-order theory of mind strategies in their repertoire, do they predominantly use a zero-order theory of mind strategy or a first-order theory of mind strategy when they fail in second-order false belief tasks?

After knowing the level of theory of mind strategies of 5-year-olds, the following research questions arise:

2. How do 5-year-olds revise their wrong theory of mind reasoning strategy to the correct second-order theory of mind reasoning strategy over time?
3. Which types of feedback help 5-year-olds to revise their wrong strategies to the correct second-order theory of mind strategy in second-order false belief tasks?

In the previous research questions, we disregard the possible roles of language and executive functions in children’s development of second-order false belief reasoning. Considering the previous literature showing that language and executive functions have a role in children’s development of theory of mind, we also investigated the following research question:

4. Does working memory or syntactic recursion predict children’s second-order false belief reasoning?

Since first-order false belief reasoning is necessary in second-order false belief reasoning, we also investigated contributing factors of children’s development of first-order false belief reasoning. Considering that children do not learn first-order false belief reasoning by practicing false belief tasks, we aimed to investigate what kind of prior cognitive skills help children to pass the first-order false belief tasks. More specifically, we focused on the following research questions for the role of children’s prior executive functioning skills:

5. What might be the role of working memory strategies in children’s development of first-order false belief reasoning?
6. What is the common mechanism of 3-year-old children’s performance on the cognitive control tasks and the first-order false belief tasks?

1.6. Methodologies

In order to investigate children's development of second-order theory of mind, this dissertation, for the first time, combines empirical experiments with computational cognitive modeling. Along with studying children's development empirically, the computational cognitive modeling approach is a powerful method to provide insight into the underlying processes of human cognition. Figure 1.4 depicts an overview of the methodology that is used in this dissertation.

Ideally, computational cognitive modelers review the previous theories and construct models based on the available theories. Constructing models gives modelers the opportunity to implement cognitive processes and cognitive concepts with precision instead of using the concepts without teasing apart their components (e.g., response inhibition, memory inhibition, goal inhibition). After implementing a computational cognitive model, modelers run simulations and the simulation result brings out new predictions. These predictions can be tested empirically. If the empirical results fit the model's predictions, the theory is revised based on the model's assumptions; otherwise, the model-prediction-experiment cycle is repeated again.

However, sometimes the literature has some empirical data that are explained by opposing theories. In these cases, computational cognitive models are constructed based on the available data and this helps to differentiate the theories more precisely and to bring out new predictions.

In particular, implementing computational cognitive models in cognitive architectures has further advantages. First, cognitive architectures are not just software for constructing cognitive models. Instead, they reflect *unified* theories of cognition (Newell, 1973; 1990) in which a wide variety of tasks can be implemented under the same architecture. Second, cognitive architectures have some parameters that are set to a default value based on previous psychological experiments to simulate average human performance. For example, it takes 200 milliseconds to press a button on the keyboard once a decision has been made and the finger is ready to press it. Therefore, there are fewer degrees of freedom in terms of the parameters of the models. Third, a single model can allow modelers to make predictions about different modalities, such as reaction times, accuracy, eye-movements and Blood Oxygenation Level Dependent signal (BOLD).

In this dissertation, I constructed computational cognitive models by using two different cognitive architectures⁵, namely ACT-R and PRIMs. Below, I give a brief overview about the relevant mechanisms of these architectures.

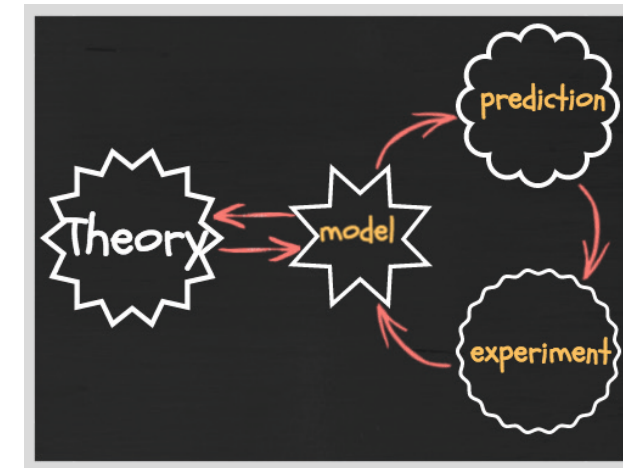


Figure 1.4. Overview of the methodology

1.6.1. ACT-R

Adaptive Control of Thought–Rational (ACT-R; Anderson, 2007) is a hybrid symbolic/sub-symbolic cognitive architecture. ACT-R consists of several modules, each associated with a specific region in the brain (see Figure 1.5).

Similar to the participants in behavioral experiments, ACT-R can “see” or “hear” the presented stimuli through its visual module, retrieve a chunk of information from its declarative module, hold the retrieved chunk in its “mind” temporarily through the problem state, change the current goal to another one through the goal state buffer, and finally, give a “verbal” or “motor” response through its manual module.

Knowledge is represented in two different memory systems in ACT-R. While the declarative memory represents factual knowledge in the form of chunks (i.e., “The capital of France is Paris”), procedural knowledge (i.e., how to ride a bicycle) is represented by the production rules in the form of IF-THEN rules. The procedural knowledge and the factual knowledge interact when production rules retrieve a chunk from the declarative memory.

At any time, the central pattern matcher checks which IF parts of the production rules match the current goal of the model, and if multiple production rules match the current goal, then the rule that has the highest utility value is executed. The utility value is calculated from estimates of the cost and probability of reaching the goal if that production rule is chosen. Noise is also added to the expected utility of a production rule, making production rule selection stochastic.

⁵ See Laird (2012), Meyer & Kieras (1997a; 1997b), Newell (1990) and Rosenbloom (2013) for other cognitive architectures.

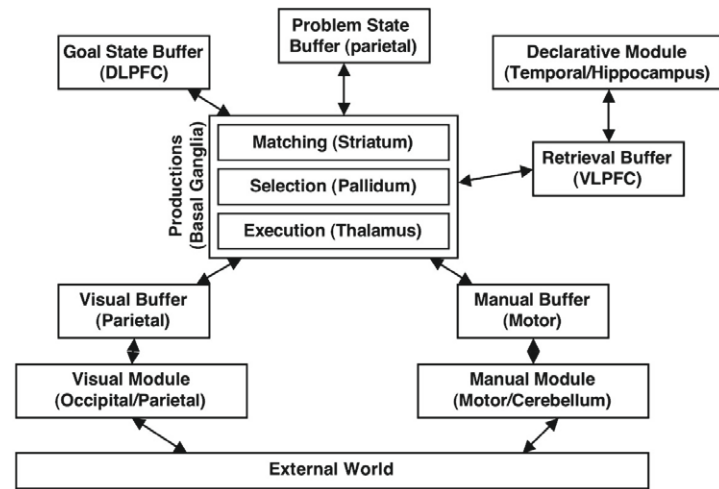


Figure 1.5. The cognitive architecture ACT-R. Reprinted from "The Nature and Transfer of Cognitive Skills," by N. A. Taatgen, 2013, *Psychological Review*, 120 (3), p. 443. Copyright 2013 by the American Psychological Association.

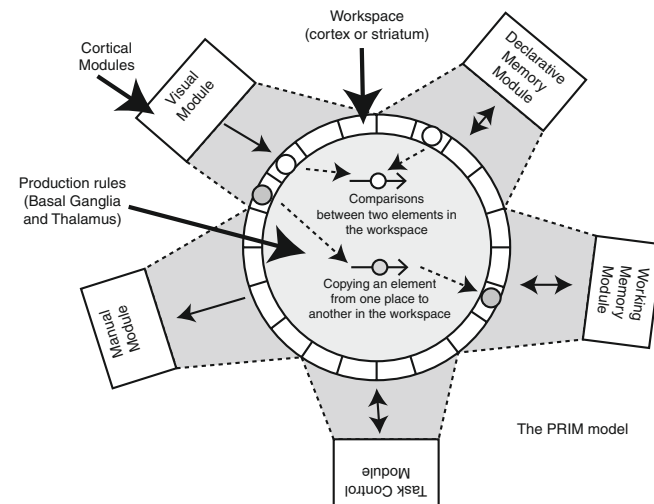


Figure 1.6. The cognitive architecture PRIMs. Reprinted from "The Nature and Transfer of Cognitive Skills," by N. A. Taatgen, 2013, *Psychological Review*, 120 (3), p. 443. Copyright 2013 by the American Psychological Association.

When a production rule is successfully executed, the central pattern matcher checks again for production rules that match the current goal. Thus, cognition unfolds as a succession of production rule executions.

Over the years, many ACT-R models have been constructed in different domains, such as perception and attention, language processing, problem solving and decision making, learning and memory, education and cognitive neuroscience⁶.

1.6.2. PRIMs

Different from ACT-R, the *primitive information processing elements* theory (PRIMs; Taatgen, 2013) has been constructed specifically as a theory of skill acquisition and transfer of skills. It has been implemented in the cognitive architecture PRIMs (see Figure 1.6).

The cognitive architecture PRIMs adopts the mechanisms of the declarative memory of ACT-R. However, in addition to chunks of factual information, the PRIMs architecture has operators and goals in declarative memory. Operators, similar to production rules in ACT-R, are in the form of IF-THEN rules (condition-action). The PRIMs architecture breaks down the complex production rules of ACT-R, which represent procedural knowledge (i.e., how to drive a car), into a fixed number of smallest possible elements, named PRIMs. PRIMs only move, compare or copy information between modules (i.e., declarative, visual, and motor modules) independent from the content of the information. Operators combine these PRIMs together to perform a task. However, unlike production rules, operators, just like other chunks in declarative memory, have base-level activations and associative strengths.

In the PRIMs architecture, there is no hard connection between goals and operators. Current goals of the model activate operators to achieve those goals. If an operator is successfully used to complete a goal, the strength of association between the goal and the operator increases.

According to the PRIMs theory, transfer of skills occurs either based on the transfer of the task-general sequences of PRIMs or based on training a particular strategy, which is represented by operators. We provide a more detailed explanation of the transfer of skills based on the transfer of the task-general sequences of PRIMs in Chapter 5 and based on training a particular strategy in Chapter 6.

Taatgen (2013) showed the predictive power of the PRIMs architecture by modeling a variety of transfer experiments such as text editing (Singley & Anderson, 1985), arithmetic (Elio, 1986), and cognitive control (Chein & Morrison, 2010).

6 See <http://act-r.psy.cmu.edu/publication/> for the list of publications.

1.7. Overview of the Dissertation

As I mentioned in Section 1.5, the main goal of this dissertation is to provide a plausible explanation to the question why children cannot pass second-order false belief tasks once they are able to pass first-order false belief tasks. We will achieve this by breaking up the components of children's development of second-order theory of mind.

One of the components we investigate is children's strategy selection and revision in second-order false belief tasks. Chapter 2 is devoted to answer the research questions: Do 5-year-olds who fail in second-order false belief tasks predominantly use zero-order theory of mind or first-order theory of mind reasoning? And how do they revise their wrong theory of mind reasoning strategy to the correct second-order theory of mind reasoning strategy over time? To this end, in Chapter 2, we present two computational cognitive models in which two possible learning mechanisms of decision making are implemented in the ACT-R cognitive architecture, namely reinforcement learning and instance-based learning. Subsequently, we present our cross-sectional study with 5-year-old children, which is conducted to test the reinforcement learning and the instance based-learning models' different predictions about five-year-olds' wrong answers to second-order false belief questions.

The second prediction of the reinforcement learning and instance-based learning models is related to the role of different types of feedback in children's development of second-order false belief reasoning. Different from the first-order theory of mind literature, both of the models predict that it is possible to accelerate 5-year-olds' second-order false belief reasoning by providing feedback "Correct/Wrong" without further explanations. However, unlike the reinforcement learning model, the instance-based learning model predicts that if feedback with further explanations is provided, children's performance would increase more. In Chapter 3, we test these predictions by training children with many different second-order false belief stories in four different conditions, namely feedback with explanation, feedback without explanation, no feedback and an active control condition in which theory of mind reasoning played no role.

Until Chapter 4, the assumption was that children do not have major problems in their executive functioning abilities to attribute second-order false beliefs and with the linguistic structure of the second-order false belief questions. In Chapter 4, we investigate two important components of children's development of second-order theory of mind, namely working memory and language. More specifically, we focus on the relationship between simple and complex working memory strategies and children's second-order false belief reasoning as well as the relationship between syntactic recursion and second-order false belief

reasoning. Furthermore, we investigate whether syntactic recursion or working memory is the best predictor of children's second-order false belief reasoning. Finally, we invoke the *serial processing bottleneck* hypothesis to propose a procedural account for the role of complex working memory strategies in second-order false belief reasoning.

Since first-order false belief reasoning is necessary in second-order false belief reasoning and since, in Chapter 4, we found that the main predictor of second-order false belief reasoning is working memory, in Chapter 5 and Chapter 6, we also investigate the role of executive functions in children's development of first-order false belief reasoning. In Chapter 5, we present a PRIMs model in order to investigate the question: How do simple and complex working memory strategies help children to pass first-order false belief tasks? In Chapter 6, we present another computational cognitive model of transfer of skills, which is implemented by using the PRIMs cognitive architecture. Our computational modeling approach provides a procedural account for the existing experimental data showing that there is a mutual transfer between children's cognitive control and first-order false belief reasoning, meaning that training 3-year-old children with a cognitive control task helps them to pass the first-order false belief task and vice versa.

Finally, in Chapter 7, I discuss our findings and give pointers to possible future research to have a better understanding of children's development of second-order theory of mind.

Chapter 2: Five-year-olds' Systematic Errors in Second-order False Belief Tasks Are Due to First-order Theory of Mind Strategy Selection: A Computational Modeling Study

*In which we investigate children's strategy
selection in second-order false belief tasks by
constructing computational cognitive models
using the cognitive architecture ACT-R.*



This chapter was previously published as:

Arslan, B., Taatgen, N. A., & Verbrugge, R. (2017a). Five-year-olds' systematic errors in second-order false belief tasks are due to first-order theory of mind strategy selection: A computational modeling study. *Frontiers in Psychology*, 8:275. doi: 10.3389/fpsyg.2017.00275

Abstract

The focus of studies on second-order false belief reasoning generally was on investigating the roles of executive functions and language with correlational studies. Different from those studies, we focus on the question how 5-year-olds select and revise reasoning strategies in second-order false belief tasks by constructing two computational cognitive models of this process: an instance-based learning model and a reinforcement learning model. Unlike the reinforcement learning model, the instance-based learning model predicted that children who fail second-order false belief tasks would give answers based on first-order theory of mind reasoning as opposed to zero-order reasoning. This prediction was confirmed with an empirical study that we conducted with 72 five- to six-year-old children. The results showed that 17% of the answers were correct and 83% of the answers were wrong. In line with our prediction, 65% of the wrong answers were based on a first-order theory of mind strategy, while only 29% of them were based on a zero-order strategy (the remaining 6% of subjects did not provide any answer). Based on our instance-based learning model, we propose that when children get feedback “Wrong”, they explicitly revise their strategy to a higher level instead of implicitly selecting one of the available theory of mind strategies. Moreover, we predict that children’s failures are due to lack of experience and that with exposure to second-order false belief reasoning, children can revise their wrong first-order reasoning strategy to a correct second-order reasoning strategy.

Keywords: second-order false belief reasoning, theory of mind, instance-based learning, reinforcement learning, computational cognitive modeling, ACT-R.

2.1. Introduction

The ability to understand that other people have mental states, such as desires, beliefs, knowledge and intentions, which can be different from one’s own, is called theory of mind (ToM; Premack & Woodruff, 1978). Many studies have shown that children who are younger than four have problems to pass verbal tasks in which they are expected to predict or explain another agent’s behavior in terms of the agent’s mental states, such as false beliefs (Wellman et al., 2001; see Onishi & Baillargeon, 2005 for an example of a non-verbal false belief task). In our daily lives, we do not only take the perspective of another agent (first-order ToM) but also use this ToM recursively by taking the perspective of an agent who is taking the perspective of another agent. For example, if David says, “Mary (falsely) believes that John knows that the chocolate is in the drawer”, he is applying second-order ToM by attributing a mental state to Mary who is attributing another mental state to John. While children start to pass verbal first-order ToM tasks around the age of four, it takes them a further one to three years to pass second-order ToM tasks (Perner & Wimmer, 1985; Sullivan et al., 1994; for a review, see Miller, 2009; 2012). Why can children not pass second-order ToM tasks once they are able to pass first-order ToM tasks? The central focus of this study is to provide a procedural account by constructing computational cognitive models⁷ to answer this question.

Many studies have shown that children who are younger than four make systematic errors in verbal first-order false belief tasks (Wellman et al., 2001). A prototype of verbal first-order false belief task is as follows: “Ayla and Murat are sister and brother. They are playing in their room. Their mother comes and gives chocolate to Murat but not to Ayla, because she has been naughty. Murat eats some of his chocolate and puts the remainder into the drawer. He doesn’t give any chocolate to Ayla. She is upset that she doesn’t get any chocolate. After that, Murat leaves the room to help his mother. Ayla is alone in the room. Because she is upset, she decides to change the location of the chocolate. She takes the chocolate from the drawer, and puts it into the toy box. Subsequently, Murat comes to the room and says he wants to eat his chocolate”. At this point, the experimenter asks a first-order false belief question: “Where will Murat look for his chocolate?” Children who are able to give the correct answer by saying “in the drawer”, correctly attribute a false belief to Murat, because he does not know that Ayla put the chocolate into the toy box. If children do not know the answer to the first-order false belief question and simply try to guess the answer, they can randomly report one of the two locations: “drawer” or “toy box”. Interestingly, most

⁷ We will use the general term ‘model’ to refer to the term ‘computational cognitive model’ for the rest of this study.

3-year-old children do not give random answers but make systematic errors by reporting the real location of the chocolate (zero-order ToM) instead of reporting the other character's false belief (first-order ToM). This systematic error is generally called 'reality bias'⁸ (Mitchell et al., 1996).

There are two dominant explanations in the first-order ToM literature for 3-year-olds' 'reality bias'. The first explanation proposes that children do not distinguish the concept of beliefs from reality, thus children need a *conceptual change* (Wellman & Gopnik, 1994; Wellman et al., 2001). The second explanation proposes that children's systematic error is due to the fact that reality is more salient to them, thus children's failure in verbal tasks are in general due to the *complexity* of the tasks, which adds further processing demands on children's reasoning processes (Birch & Bloom, 2004; 2007; Carlson & Moses, 2001; Epley et al., 2004; Hughes, 2002;). More specifically, children automatically reason about their own perspective and in order to give an answer about another agent's perspective which is different from the reality, they should first inhibit their own perspective and then take into account the other agent's perspective and give an answer accordingly (Leslie & Polizzi, 1998; Leslie et al., 2004; Leslie et al., 2005). The debate is still on about the possible reasons of children's 'reality bias' (Baillargeon et al., 2015; Hansen, 2010; Helming et al., 2014; Lewis et al., 2012; Rubio-Fernández, 2017). In any case, it is known that most of the typically developing children around the age of 5 are able to pass first-order false belief tasks. Therefore, we can safely assume that 5-year-old children's conceptual development of reasoning about another agent's false beliefs and their executive functioning abilities to inhibit their own perspective are already well developed. This means that 5-year-olds have both efficient zero-order and first-order ToM strategies in their repertoire. Furthermore, we argue that although 5-year-olds are able to attribute second-order mental states to other agents, they are not used to answering questions that require second-order false belief attribution, which is why they need sufficient exposure to second-order false belief stories to revise their strategy.

Similar to the first-order false belief tasks, second-order false belief tasks are used to assess the continuation of children's ToM development after the age of 4. Regardless of the variations in the second-order false belief tasks (see Perner & Wimmer, 1985; Sullivan et al., 1994), they provide two critical pieces of information in addition to the first-order false belief task for which we introduced a prototype above. The first addition for the prototype story is: "While Ayla is changing the location of the chocolate, Murat passes by the window, and he sees

⁸ It is also called 'egocentrism' (Piaget, 1930) or the 'curse of knowledge' (Birch & Bloom, 2004). Although these different terms correspond to some underlying theoretical differences (see Birch & Bloom, 2004 and Mitchell et al., 1996 for these differences), we use the term 'reality bias' in this study to refer to children's systematic errors based on reality.

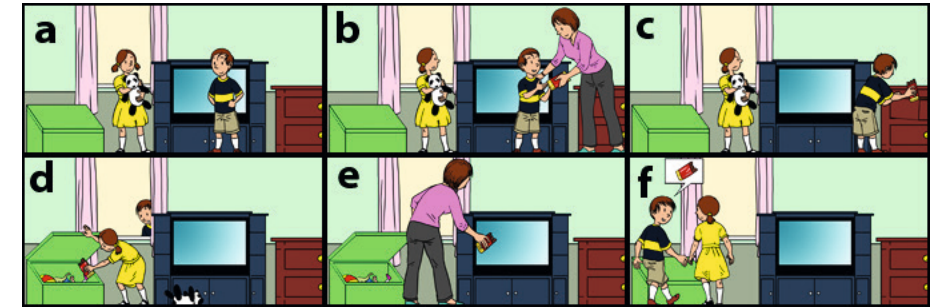


Figure 2.1. A modified version of the standard 'unexpected location' second-order false belief stories (Illustration ©Avik Kumar Maitra).

how Ayla takes the chocolate from the drawer and puts it into the toy box". The second additional aspect is: "Ayla does not notice that Murat sees her hiding the chocolate" (Figure 2.1d). Therefore, Ayla has a false belief about the location of the chocolate (i.e., Ayla *thinks that* Murat *believes that* the chocolate is in the drawer). The second-order false belief question for this prototype is as follows: "Where does Ayla think that Murat will look for the chocolate?"⁹. If children correctly attribute a false belief to Ayla, who *thinks that* Murat *believes that* the chocolate is in the drawer, they give the correct answer "drawer". Otherwise, they give the wrong answer "toy box". However, the answer "toy box" would be the correct answer to both the question "Where is the chocolate now?" (zero-order ToM), and the question "Where will Murat look for the chocolate?" (first-order ToM). That is why it is not possible to distinguish whether the wrong answer "toy box" to the second-order false belief question is due to applying a zero-order or a first-order ToM strategy.

To the best of our knowledge, there is no study that has a specific prediction together with a possible explanation about the level of ToM reasoning in children's wrong answers in second-order false belief tasks. However, a modified version of the standard second-order false belief task in which it is possible to distinguish children's level of ToM reasoning has been constructed (Hollebrandse et al., 2008). Following our prototype of the standard second-order false belief story that we mentioned above, a prototype of the modified version of the second-order false belief story has the following additional information: After telling the children that Ayla does not know that Murat saw her hiding the chocolate in the toy box, the children are informed that the mother of Ayla and Murat comes to the room when both Ayla and Murat are not there. The mother finds the chocolate in

⁹ For another type of second-order false belief question see the 'Three goals' story prototype in Section 3.2.

the toy box while she is cleaning the room, takes it out of the toy box, and puts it into the TV stand (Figure 2.1e). This modification allows us to distinguish children's zero-order ToM answers ("TV stand") from their first-order ToM answers ("toy box") for the second-order false belief question "Where does Ayla think that Murat will look for the chocolate?"

Considering our central question why children cannot pass second-order false belief tasks once they are able to pass first-order false belief tasks, a new question about strategy selection arises: Once 5-year-old children already have zero-order and first-order ToM strategies in their repertoire, do they predominantly use a zero-order ToM strategy or a first-order ToM strategy when they fail in second-order false belief tasks? There are two contradictory findings about children's systematic errors on second-order false belief tasks. Hollebrandse et al. (2008) tested 35 American-English 7-year-old children (range: 6;1 – 7;10, mean = 6;11) with a modified version of a second-order false belief task. The goal of their study was to investigate the acquisition of recursive embedding and its possible relation with recursive ToM. Their results about the second-order false belief task showed that while 58% of the answers were based on second-order ToM strategy, 32% of the answers were based on a first-order ToM strategy, and none of the answers was based on a zero-order ToM strategy. In contrast, de Villiers et al.'s (2014) preliminary results showed that 60% of five- to six-year-olds' answers were based on the zero-order ToM strategy, and only around 20% of children's answers were based on the first-order ToM strategy in the second-order false belief task. Different from those studies, our empirical study was designed to investigate children's level of wrong answers, and we had a model-based prediction about children's systematic errors in second-order false belief tasks before conducting the empirical study.

Another important question is: What do children need for revising their wrong strategy to a correct second-order ToM strategy? Analogous to the first-order ToM literature, two possible explanations have been proposed for children's development of second-order ToM: i) *conceptual change*, and ii) *complexity* (Miller, 2009, p. 751; Miller, 2012). The pure *conceptual change* explanation suggests that children's failure in the second-order ToM tasks is due to their lack of realization that mental states such as beliefs can be used recursively (e.g., "John thinks that David believes that..."). On the other hand, the pure *complexity explanation* suggests that it is the higher complexity of second-order ToM reasoning that adds further demands on working memory, as does the linguistic complexity of the stories and the questions, in comparison to first-order ToM tasks.

In order to provide a procedural account for children's ToM strategy revision, we constructed two computational cognitive models by implementing two possible learning mechanisms. The first is based on reinforcement learning (Shah,

2012; Sutton & Barto, 1998; van Rijn et al., 2003). This type of learning is based on the utilities of the rules that carry out the possible strategies. Based on feedback, a reward/punishment is propagated back in time through the rules that have been used to make the decision. This reward/punishment mechanism updates the utility of those rules and finally the model learns to apply a correct strategy.

The second model is based on instance-based learning (Gonzalez & Lebiere, 2005; Logan, 1988; Stevens et al., 2016). The central idea in instance-based learning is that decisions are based on past experiences that are stored in memory. Whenever a decision has to be made, the most active experience is retrieved from memory and used as the basis for the decision. Activation is based on history (how frequent and recent was the experience) and on similarity (how similar is the context of the past decision to the present experience). An advantage of instance-based learning is that feedback can be used to create an instance that incorporates the correct solution.

We used instance-based learning for the selection of different levels of ToM strategies (i.e., zero-order, first-order, second-order) that are stored in the declarative memory. When the model is correct in using a particular level of ToM, it will strengthen the instance related to that level, but when the model makes a mistake, it will add an instance for the next level.

Instead of adopting either the pure *conceptual change* or the pure *complexity* explanation, we argue that the following steps are followed. First, children should be aware that they can use their first-order ToM strategy recursively. Importantly, different from the reinforcement learning model, the instance-based learning model explicitly revises its strategy, therefore, it satisfies this condition. After that, children have to have efficient cognitive skills to carry out second-order ToM reasoning without mistakes. In the scope of this study, we assume that 5-year-olds have efficient cognitive skills to carry out the second-order ToM strategy. Finally, children need enough experience to determine that the second-order ToM strategy is the correct strategy to pass second-order false belief tasks (see Goodman et al., 2006 for a model of children's development of first-order false belief reasoning based on experience; and Gopnik & Wellman, 1992 for the theory that children are rational agents and that with additional evidence they revise their theories, just like scientists do).

Both the reinforcement learning model and the instance-based learning model strengthen or revise their strategies based on experience and the feedback "Correct/Wrong" without further explanation. Is it possible to assume that children get feedback "Correct/Wrong" in ToM-related tasks in their everyday life? There can be many social situations in which children get the feedback "Correct/Wrong", not from a person who gives feedback verbally, but from other consequences of a particular ToM strategy. For example, young children who are

not able to apply a first-order ToM strategy are generally unable to hide themselves properly when they are playing hide and seek¹⁰ (e.g., they hide themselves behind the curtain while their feet are visible or basically they close their eyes with their hands without hiding themselves). In this case, the feedback “Wrong” would be conferred by the fact that the seeker finds the hider immediately. Similarly, at a later stage of development, imagine a child secretly eating some of the chocolates that his mother explicitly told him not to eat. As soon as his mother comes back to the room, the child says that he does not like chocolate with nuts. His mother gets angry and tells him to go to his room and not to join them for dinner. In this case, the child was unable to use a second-order ToM strategy (i.e., *my mother should not know that I know that there are chocolates with nuts*) and although he does not get any explanations, he does get the feedback “Wrong”. Note that in this example, the child also requires other types of reasoning, such as causal reasoning, in addition to second-order ToM.

The main differences between the reinforcement learning model and the instance-based learning model derive from the way they handle the feedback when the given answer is wrong. While the reinforcement learning model punishes the strategies that lead to a wrong answer, the instance-based learning model adds an instance of another strategy. This is because while the strategy selection is implicit in the reinforcement learning model, it is explicit in the instance-based learning model. Moreover, if feedback with further explanations is provided, the instance-based learning model will more likely use a second-order ToM strategy, because it explicitly increments the level of ToM strategy to a higher ToM strategy. On the other hand, the reinforcement learning model can do nothing with the further explanations. We provide more detailed explanations for these two learning mechanisms in our models in Subsection 2.2.2. Importantly, these two models provide different predictions about children's wrong answers in second-order false belief tasks. We present those predictions in Subsection 2.2.7.3 and test them in an empirical study in Section 2.3.

In the following section, we first review the previous computational models of verbal first-order and second-order false belief reasoning. After that, we discuss the relevant mechanisms of the cognitive architecture ACT-R. Subsequently, we explain our instance-based and reinforcement learning models and their results and predictions.

¹⁰ See <https://www.youtube.com/watch?v=u03VidFILmg> for a video of young children's failure in hide and seek.

2.2. A model of second-order false belief reasoning

Along with studying children's development empirically, the modeling approach is a powerful method to provide insight into the underlying processes of children's performance (see Taatgen & Anderson, 2002 for an example on how children learn irregular English verbs without feedback; van Rij et al., 2010 for an example of the underlying processes of children's poor performance on pronoun interpretation). In particular, using cognitive architectures (e.g. ACT-R: Anderson, 2007; SOAR: Laird, 2012; SIGMA: Rosenbloom, 2013; PRIMs: Taatgen, 2013) gives us the opportunity to make specific predictions about children's accuracy, reaction times, and even the brain regions that are activated when they perform a task. These predictions can then be tested empirically.

In general, cognitive architectures have certain general assumptions about human cognition and have some parameters that are set to a default value based on previous psychological experiments to simulate average human performance. For example, it takes 200 milliseconds to press a button on the keyboard once a decision has been made and the finger is ready to press it. In addition to these general assumptions, modelers make their own specific assumptions about the tasks that they are modeling and those assumptions can be tested empirically together with the model's simulation results. Because it is always possible to make a fit to data by changing the parameters, it is preferable not to change these default values of the architecture and not to introduce new parameters unless there is a good explanation for doing so.

In this study, we use the cognitive architecture ACT-R (Anderson & Lebiere, 1998; Anderson, 2007). Before providing information about ACT-R and our models, in the following subsection, we review the previous computational cognitive models of verbal first-order and second-order false belief reasoning.

2.2.1. Previous models of false belief reasoning

Only few computational cognitive models of verbal false belief reasoning have been constructed in the literature, aiming to contribute to theoretical discussions by providing explanations. Most of those models aimed to explain children's development of first-order false belief reasoning.

Goodman et al. (2006) approach the development of first-order false belief reasoning as rational use and revision of intuitive theories, instead of focusing on children's limitations in processing information. By using Bayesian analysis, they simulate the transition from a model that represents children's reasoning from their own perspective (zero-order ToM) to another model that takes into

account another agent's perspectives (first-order ToM). Initially, the zero-order ToM model is preferred due to the Bayesian Occam's razor effect. Subsequently, based on experience with first-order false belief reasoning, the first-order ToM model becomes the preferred model thanks to its explanatory power. Bello and Cassimatis' (2006) rule-based model showed that explicit reasoning about beliefs of another agent might not be necessary in order to pass first-order false belief tasks and that it is enough to relate people to alternate states of affairs and to objects in the world. Hiatt and Trafton (2010) simulated the gradual developmental of first-order ToM by using reinforcement learning. Their models have a good match to the available gradual development data in the literature. However, they introduced additional parameters to the core cognitive architecture, namely a "selection parameter" representing increasing functionality of the brain in children's development, and a "simulation parameter" that determines the availability of rules for simulation in predicting another person's action (i.e., if the simulation parameter is 0, the model is not able to predict another's action). Therefore, the transition from zero-order ToM reasoning to first-order ToM reasoning is achieved by manipulating those parameters. More recently, Arslan et al.'s (2015b) model predicted that training children with working memory tasks might also contribute to the transition from failure to success in first-order false belief tasks.

To the best of our knowledge, there are only two computational cognitive modeling studies of second-order false belief reasoning. Wahl and Spada (2000) modeled a competent child's reasoning steps in a second-order false belief task by using a logic programming language. Their simulations predicted that explanation of a second-order false belief attribution is more complex than its prediction. They validated their model-based prediction with an empirical study with children between the ages 6 to 10. For future research, they suggested to use a cognitive architecture such as ACT-R to simulate children's incorrect answers. Recently, similar to their first-order false belief reasoning model, Hiatt and Trafton (2015) simulated the gradual developmental of second-order ToM by using reinforcement learning. Again, their model had a good match to the available data for the developmental trajectory of second-order ToM. However, they kept the "selection parameter" and "simulation parameter" that they introduced to the default parameters of ACT-R and they did not provide any specific predictions that can be tested empirically.

Different from the available second-order ToM models, we set the following criteria when constructing our models:

- i. The models should simulate children's transitions from incorrect to correct answers in second-order false belief tasks;

- ii. The transition to second-order reasoning should naturally emerge from the simulation, and should not be controlled by mechanisms that are not part of the cognitive architecture (i.e., ACT-R);
- iii. The models should provide predictions that can be tested empirically, before conducting a behavioral experiment.

Considering the above-mentioned criteria, we explore two different learning mechanisms of ACT-R, namely instance-based learning and reinforcement learning, to be able to compare their predictions.

2.2.2. The relevant mechanisms of the cognitive architecture ACT-R

ACT-R is a hybrid symbolic/sub-symbolic production-based cognitive architecture (see Anderson, 2007 for a detailed overview). Knowledge is represented in two different memory systems in ACT-R.

While the declarative memory represents the factual knowledge in the form of chunks (i.e., "The capital of France is Paris"), procedural knowledge (i.e., how to ride a bicycle) is represented by the production rules in the form of IF-THEN rules. The procedural knowledge and the factual knowledge interact when production rules retrieve a chunk from the declarative memory. At any time, the central pattern matcher checks the IF part of the production rules that match the current goal of the model, and if multiple production rules match the current goal, the rule that has the highest utility value is executed. The utility value is calculated from estimates of the cost and probability of reaching the goal if that production rule is chosen. Noise is also added to the expected utility of a production rule, making production rule selection stochastic. When a production rule is successfully executed, the central pattern matcher checks again for production rules that match the current goal. Thus, cognition unfolds as a succession of production rule executions.

For models of learning in decision making, there are two categories of solutions in ACT-R: i) instance-based learning, ii) reinforcement learning. *Instance-based learning* occurs by adding new chunks to the declarative memory. If an identical chunk is already in memory, the new chunk is merged with the previous identical chunk and their activation values are combined. Each chunk is associated with an activation value that represents the usefulness of that chunk. The activation value of a chunk depends on its base-level activation (B) and on activation sources originating in the model's context. The base-level activation is determined by the frequency and recency of a chunk's use together with a

noise value (Anderson & Schooler, 2000). A chunk will be retrieved if its activation value is higher than a retrieval threshold, which is assigned by the modeler. While a chunk's activation value increases each time it is retrieved, its activation value will decay over time when it is not retrieved. Depending on the type of the request from declarative memory, the chunk with the highest activation value is retrieved. The optimized learning equation which is used in the instance-based learning model to calculate the learning of base-level activation for a chunk i is as follows:

$$B_i = \ln(n/(1-d)) - d * \ln(L)$$

Here, n is the number of presentations of chunk i , L is the lifetime of chunk i (the time since its creation), and d is the decay parameter.

In ACT-R, *reinforcement learning* occurs when the utilities (U) that are attached to production rules are updated based on experience (Taatgen et al., 2006). A strategy (i.e., zero-order, first-order, second-order) that has the highest probability of success is used more often. Utilities can be updated based on rewards (R). Rewards can be associated with specific strategies, which are implemented by production rules. The reward is propagated back to all the previous production rules that are between the current reward and the previous reward. The reward that is propagated back is calculated with the assigned reward value minus the time passed since the execution of the related production rule, meaning that more distant production rules receive less reward. If the assigned reward is zero, the production rules related to the execution of a production rule that is associated with the reward will receive negative reward (punishment). Based on these mechanisms, a model learns to apply the best strategy for a given task. The utility learning equation which is used in the reinforcement learning model is as follows:

$$U_i(n) = U_i(n-1) + \alpha[R_i(n) - U_i(n-1)]$$

Here, $U_i(n-1)$ is the utility of a production i after its $n-1^{\text{st}}$ application, $R_i(n)$ is the reward the production receives for its n^{th} application and α is the learning rate.

In the following two sections, we explain the details of how the instance-based learning and reinforcement learning model select theory of mind strategies based on experience. Subsequently, we explain the general assumptions and the reasoning steps in both models.

2.2.3. How the instance-based learning model goes through transitions

The assumption of the instance-based learning model is that possible strategies to apply different levels of ToM reasoning in the second-order false belief task (i.e., zero-order, first-order, second-order) are represented as chunks in declarative memory. The model uses these to select its strategy at the start of a problem: It will retrieve the strategy with the highest activation, after which production rules carry out that strategy. Based on the success, the model will either strengthen a successful strategy chunk, or will add or strengthen an alternative strategy if the current one failed. Our instance-based learning model uses the same mechanism for strategy selection as in Meijering et al.'s (2014) ACT-R model that shows adults' strategy selection in a ToM game. The core idea of their model was that people in general use ToM strategies that are "as simple as possible, as complex as necessary" so as to deal with the high cognitive demands of a task.

The instance-based learning model starts with only a single strategy, which is stored in declarative memory as a chunk: the zero-order ToM strategy. Similar to young children's daily life experiences, the zero-order ToM strategy chunk's base level activation is set to a high value to represent that the model has a lot of experience in using this strategy. In line with this simplistic zero-order ToM strategy that is based on the real location of the object, the model gives the answer "TV stand" (see Figure 2.1) to the second-order false belief question ("Where does Ayla think that Murat will look for the chocolate?"). However, as this is not the correct answer to the second-order false belief question (drawer), the model gets the feedback "Wrong" without any further explanation. This stage of the model in which the zero-order ToM strategy seems to be more salient than the first-order ToM strategy represents children who are able to attribute first-order false beliefs but are lacking experience in applying the first-order ToM strategy.

Given this feedback, the model increments the reasoning strategy just used (zero-order) one level up and enters a new strategy chunk in declarative memory: a chunk that represents the first-order ToM strategy, in which the former (zero-order) strategy is now attributed to Murat. This makes it a first-order ToM strategy because this time the model gives an answer based on what the reality is (zero-order) from Murat's perspective (first-order). Because the model has more experience with the zero-order ToM strategy, the activation of the zero-order ToM strategy chunk at first is higher than the recently added first-order ToM strategy chunk. This causes the model to retrieve the zero-order ToM strategy chunk instead of the first-order ToM strategy chunk in the next few repetitions of the task. Thus, the model still gives an answer to the second-order false belief question based on zero-order reasoning. Nevertheless, each time that the model gets the

negative feedback “Wrong”, it creates a first-order ToM strategy chunk. As the identical chunks are merged in the declarative memory, the first-order ToM strategy chunk's activation value increases.

When the activation value of the first-order ToM strategy chunk is high enough for its successful retrieval, the model gives an answer to the second-order false belief question based on first-order reasoning (toy box). Again, this is not a correct answer to the second-order false belief question (drawer). After the model gets the feedback “Wrong”, it again increments the first-order strategy by attributing a first-order ToM strategy to another agent (Ayla), which makes it a second-order ToM strategy, because this time the model gives an answer based on what Murat *thinks* (first-order) from Ayla's perspective (second-order). This second-order strategy gives the correct answer (drawer). Given the positive feedback “Correct”, the second-order ToM strategy is further strengthened and finally becomes stable. In theory, there is no limitation on the level of strategy chunks. Nevertheless, in practice there is no need to use a very high level of reasoning (Meijering et al., 2014), and even if one tries to apply more than third-order or fourth-order ToM reasoning, it will be very hard to apply that strategy due to memory limitations (see Kinderman et al., 1998 and Stiller & Dunbar, 2007 for adults' limitations in higher levels of ToM reasoning).

2.2.4. How the reinforcement learning model goes through transitions

Unlike our instance-based learning model in which the reasoning strategy chunks (i.e., zero-order, first-order, second-order) are added to the declarative memory over the repetition of the task, in the reinforcement learning model the reasoning strategies are implemented with production rules. Therefore, the model selects one of these strategies based on their utilities.

Similar to the zero-order ToM strategy chunk's relatively high base-level activation in the instance-based learning model, the utility of the production rule of the zero-order ToM strategy is arbitrarily set to a much higher value (100) than the production rules that represent the first-order (25) and second-order (5) ToM strategies. Thus, initially the reinforcement learning model gives zero-order answers. The relativity of those values is the assumption that the model has a lot of experience with the zero-order ToM strategy, and more experience with the first-order ToM strategy than the second-order ToM strategy, based on children's development.

After the reinforcement learning model gives the zero-order answer (TV stand), it gets the feedback “Wrong”. Based on this feedback, the zero-order ToM

strategy production rule gets zero reward. As explained in Section 2.2.2, this mechanism decreases the utility of the zero-order ToM strategy production rule. The first-order ToM strategy production rules are executed when the utility of the zero-order ToM strategy decreases enough (to around 25). After selection of a first-order ToM strategy, the model again gets zero reward. This reward is propagated back through the other production rules of the first-order ToM strategy up to the production rule that gives the zero-order answer. Finally, when the model is able to execute the second-order strategy and to give the correct answer (drawer), it gets a higher reward (20). Therefore, the second-order ToM strategy becomes the dominant strategy. Importantly, as we discussed in the Introduction, the selection of ToM strategies is purely based on a utility mechanism, thus it is implicit compared to the explicit ToM strategy of the instance-based model.

2.2.5. General assumptions and reasoning steps in both models

Even though our models are not dependent on the particular features of a specific second-order false belief task, we modeled children's reasoning in the prototype of a modified version of the standard second-order false belief that we explained in the introduction (see Figure 2.1). One of the assumptions of our models is that the models already heard the second-order false belief story and are ready to answer the second-order false belief question “Where does Ayla think that Murat will look for the chocolate?” Thus, the story facts are already in the models' declarative memory. The models do not store the entire story in their declarative memory but just the facts that are related to answering the second-order false belief question. Table 1 presents the verbal representation of those story facts¹¹. As can be seen from Table 1, each story fact is associated with a specific time, meaning that the model knows which events happened after, before, or at the same time as a certain other event. Unlike the reinforcement learning model, the instance-based learning model starts with a zero-order ToM strategy chunk in declarative memory in addition to the story facts.

Both models have the following task-independent knowledge to answer the second-order false belief question (see Stenning & van Lambalgen, 2008 for an example formalization of a first-order false belief task by using similar task-independent knowledge): i) The location of an object changes by an action towards that object; ii) ‘Seeing leads to knowing’, which is acquired by children around the age of 3 (Pratt & Bryant, 1990); iii) People search for objects at the

¹¹ For the actual representations in ACT-R, see <https://figshare.com/s/7c3146ad85b3e7a57cd4>.

Table 2.1. The representations of story facts that are initially in declarative memory before the model starts to reason for the second-order false belief question.

"Murat put the chocolate into the drawer at time t1"	(Figure 2.1c)
"Ayla put the chocolate into the toy box at time t2"	(Figure 2.1d)
"Murat saw Ayla at time t2"	(Figure 2.1d)
"Ayla did not see Murat at time t2"	(Figure 2.1d)
"The mother put the chocolate into the TV stand at time t3"	(Figure 2.1e)

location where they have last seen them unless they are informed that there is a change in the location of the object; iv) Other people reason ‘like me’. For instance, based on the task-independent knowledge (ii), both models can infer that Murat knows that the chocolate is in the toy box once the story fact “Murat saw Ayla at time 2” (Table 1, row 3) has been retrieved.

Table 2.2 shows the steps that have been implemented to give an answer for the second-order false belief question in the instance-based model and the reinforcement learning model. As can be seen from Table 2.2, both models always use the same set of production rules in the first two steps, which represent reasoning about reality. This feature of the models reflects the usual process of a person’s reasoning from his/her own point of view (Epley et al., 2004). Although the instance-based and reinforcement learning models have different learning mechanisms and different underlying assumptions for the selection of the reasoning strategies, the general idea for both models is that they reason about another agent as if the other model is reasoning “like me”, and use this “like me” strategy recursively. Note that we implemented the models to answer the second-order false belief question, therefore, the second-order ToM strategy becomes stable over repetition. However, when the models hear the first-order false belief question “Where will Murat look for the chocolate?”, they will use a first-order strategy instead of a second-order reasoning strategy if the activation of the first-order strategy is higher than the zero-order strategy in the instance-based learning model and if the utility of the first-order reasoning strategy is higher than that of the zero-order strategy in the reinforcement learning model.

In more detail, the k^{th} -order reasoning production rules shared by both of the models are as follows: If the *zero-order strategy* is retrieved, give an answer based on the location slot of the chunk that has been retrieved previously. If the *first-order strategy* is retrieved, check whether Murat saw the object in that location or not. If Murat saw the object in that location, give an answer based on the location slot of the chunk, otherwise retrieve a chunk in which Murat saw the object previously

Table 2.2. The steps that are implemented to give an answer for the second-order false belief question for the instance-based and the reinforcement learning models

Instance-based learning model	Reinforcement learning model
1. Retrieve a story fact that has an action verb in its slots.	1. Retrieve a story fact that has an action verb in its slots.
2. Check the time slot of the retrieved story fact and if it is not the latest fact, request the latest one.	2. Check the time slot of the retrieved story fact and if it is not the latest fact, request the latest one.
3. Request a retrieval of one of the strategy chunks from declarative memory.	3. If the production rule that represents the zero-order strategy has the highest utility, give an answer based on the location slot of the chunk that is retrieved in the second step. If the production rule that represents k^{th} -order strategy ($0 < k \leq 2$) has the highest utility, apply that strategy to give an answer by reasoning as if that person employs $(k-1)^{th}$ -order reasoning.
4. If the zero-order strategy is retrieved, give an answer based on the location slot of the chunk that is retrieved in the second step. If the k^{th} -order strategy ($0 < k \leq 2$) is retrieved, determine whose knowledge the question is about and give the answer by reasoning as if that person employs $(k-1)^{th}$ -order reasoning.	4. Based on the feedback (i.e. Correct/Wrong), strengthen the successful strategy chunk, or will add or strengthen an alternative strategy if the current one failed.
5. Based on the feedback (i.e. Correct/Wrong), strengthen the successful strategy chunk, or will add or strengthen an alternative strategy if the current one failed.	

and give an answer based on the location slot of that chunk. If the *second-order strategy* is retrieved, repeat the procedure of the first-order strategy, however, this time, instead of giving the answer from Murat’s perspective, check whether Ayla saw Murat at that time. If Ayla did not see Murat, then retrieve a chunk in which Murat put the object and give an answer based on the location slot of that chunk¹².

Therefore, the generalized explanation of this procedure can be summarized as follows: If the k^{th} -order strategy ($0 < k \leq 2$) is retrieved, determine whose knowledge the question is about and give the answer by reasoning as if that person employs $(k-1)^{th}$ -order reasoning.

2.2.6. Parameters

Following the criteria that we stated in Section 2.2.1, we did not introduce any new parameters in addition to ACT-R’s own parameters. Moreover, all the parameters

¹² It works similarly for the other possible first-order and second-order ToM questions in which Ayla and Murat appear (i.e., “Where will Ayla look for the chocolate?” and “Where does Murat think that Ayla will look for the chocolate?”).

were set to their default values, except the retrieval threshold and the instantaneous noise parameters for the instance-based learning model and the utility noise parameter for the reinforcement learning model (there are no default values in ACT-R for those parameters). As previous empirical studies showed that children mostly give correct answers for the control questions (Flobbe et al., 2008; Hollebrandse et al., 2008), the retrieval threshold was set to an arbitrary low value (-5), so that the model is always able to retrieve the story facts. Thus, our models' failure in the second-order false belief task is not due to forgetting some of the story facts but due to inappropriate strategy selection.

For the reinforcement learning model, we turned the utility learning parameter on. Similar to activations, noise is added to utilities. Noise is controlled by the utility noise parameter, which is set to 3.

2.2.7. Model results and predictions

In this subsection, we present the results and predictions of the instance-based model and the reinforcement learning model.

2.2.7.1. Instance-based learning model results

To show the developmental transitions from zero-order to second-order reasoning, we ran the model 100 times per 'virtual child', indicating that one child learns to apply second-order reasoning over time by gaining experience. To average the results across 100 children, we made 100 repetitions of the second-order false belief task for each 'virtual child'. Thus, we ran the model 10,000 times in total. For each 'virtual child', the initial activation of the zero-order reasoning chunk was set to 6, indicating that children have a lot of experience with zero-order reasoning. Figure 2.2a shows the proportion of the levels of reasoning the model applies, and Figure 2.2b shows the activation values of the strategy chunks over time.

In Figure 2.2a, around the 12th repetition, the model uses the first-order strategy (60%) more than the zero-order strategy (40%). Around the 26th repetition, the model uses both second-order (50%) and first-order (50%) reasoning with an equal chance, and around the 40th repetition the model uses second-order reasoning (80%) much more than first-order reasoning (20%). Finally, around the 50th repetition, the second-order reasoning strategy becomes stable (100%).

As we explained in Section 2.2.3, the transitions in the strategy chunks are based on the activation of those chunks. In Figure 2.2b, around the 10th repetition,

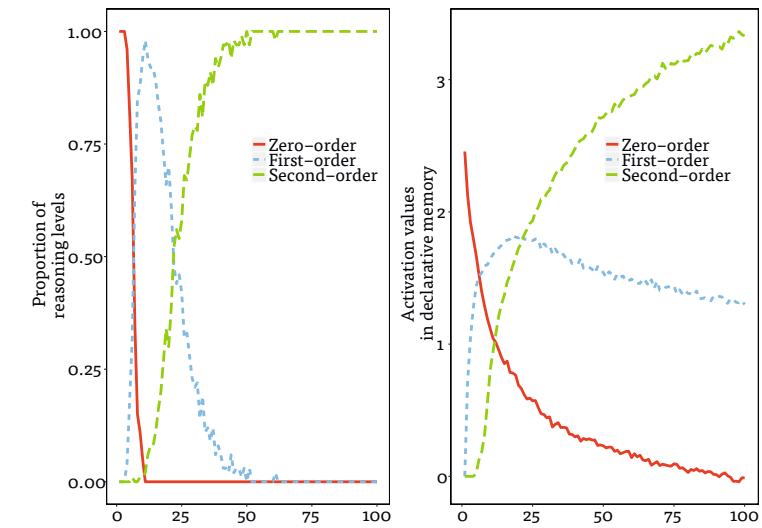


Figure 2.2. (a) Proportions of the reasoning level that the instance-based learning model applies, and (b) the activation values of the reasoning level strategy chunks, plotted as a function of number of repetitions.

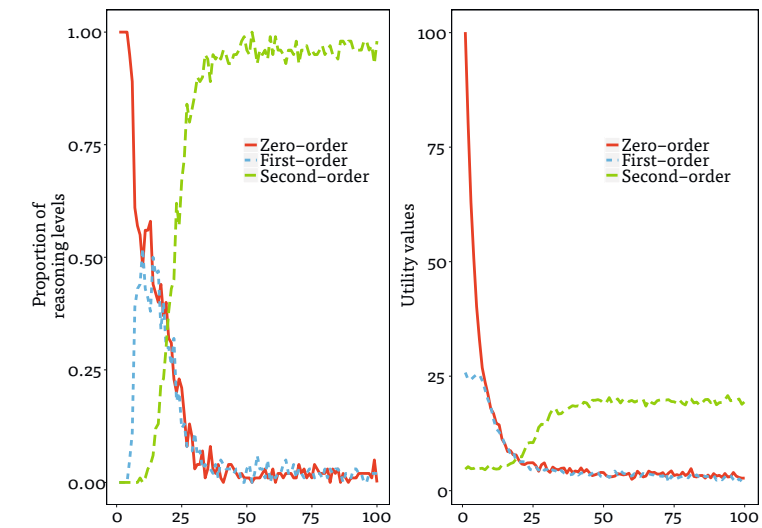


Figure 2.3. (a) Proportions of the reasoning level that the reinforcement learning model applies, and (b) the utility values of the strategies, plotted as a function of number of repetitions.

the first-order ToM strategy chunk's activation becomes higher than that of the zero-order ToM strategy chunk which leads the model to apply first-order ToM instead of the zero-order ToM. Finally, around the 26th repetition, the activation of the second-order ToM strategy chunk's activation becomes higher than that of the other strategies, so that the model makes a second-order belief attribution to Ayla.

2.2.7.2. Reinforcement learning model results

Similar to the instance-based learning model, we ran the reinforcement learning model 10,000 times in total to average the results across 100 'virtual children' repeating the second-order false belief task 100 times each. Figure 2.3a shows the proportion of the levels of reasoning that the reinforcement learning model applies, and Figure 2.3b shows the utility values of the strategies.

Different from the instance-based learning model's results (see Figure 2.3a), the reinforcement learning model does not go through the transitions in a step-wise fashion. Until around the 10th repetition, the model uses a zero-order strategy and a first-order strategy randomly (50%/50%), and does not use the second-order strategy. Before the model starts to use the second-order strategy more often (60%) than the other two strategies (around 30th repetition), it uses both the zero-order and first-order strategies, and not necessarily the first-order strategy more often than the zero-order one. Finally, around the 50th repetition, the second-order reasoning strategy becomes stable (100%).

2.2.7.3. Comparing the predictions of the two models

1. The first predictions of the instance-based and reinforcement learning models are related to children's errors in second-order false belief tasks. Following the pattern in Figure 2.2a, the *instance-based learning* model predicts that children who do not have enough experience with second-order reasoning give first-order answers to the second-order false belief question. On the other hand, following the pattern in Figure 2.3a, once the *reinforcement learning* model is able to execute the first-order ToM strategy, it selects between zero-order and first-order ToM strategies randomly, on the basis of noise. Thus, the reinforcement learning model does not predict that children's wrong answers would most of the time be based on the first-order reasoning strategy¹³.

¹³ Note that one can argue that the predictions of the reinforcement learning model may be changed by just adjusting the initial utilities of the strategies or the added noise parameters. While changing the initial utility values

2. The second predictions of both models are related to learning second-order false belief reasoning over time based on the given feedback. Both of the models predict that children who have enough experience with first-order ToM reasoning but not with *second-order* ToM reasoning can learn to apply second-order ToM without any need to have further explanations of why their answer is wrong. This prediction contrasts with previous findings showing that 4-year-old children's performance on *first-order* false belief tasks cannot be improved when they are trained on false belief tasks with feedback without giving detailed explanations (Clements et al., 2000).
3. Although both models predict that training children with feedback "Wrong" is sufficient to accelerate their development of second-order false belief reasoning, the *instance-based* learning model provides an additional underlying prediction. Because the instance-based learning model explicitly increments its wrong first-order ToM strategy to the correct second-order ToM strategy, if the model would receive feedback *together with further explanations* (not only "Wrong"), the odds of selecting the correct strategy would increase. In contrast, providing feedback *with further explanations* does not provide any useful additional information for the reinforcement learning model.

2.3. Experimental validation of the instance-based learning model

In this section, we present the experimental validation of our instance-based learning model's first prediction, which proposes that 5-year-old children will give first-order ToM answers in the second-order false belief task.

2.3.1. Participants

In order to test our model-based predictions related to children's wrong answers, we analyzed the cross-sectional data (pre-test) of a larger training study that includes a sample of 79 Dutch 5- to 6-year-old children (38 female, $M_{\text{age}} = 5.7$ years, $SE = 0.04$, range: 5.0 – 6.8 years). All children were recruited from a primary school in Groningen, the Netherlands from predominantly upper-middle-class families. The children were tested individually in their school in a separate room.

or the noise values would change the exact curves in Figure 2.3A, both manipulations would not change the reinforcement learning model's prediction unless the second-order ToM strategy has higher utility than the zero-order and first-order ToM strategies, which is theoretically not plausible (see S1 Materials for examples of the models' results with different utility values and different noise values).

Approval and parental consent was obtained in accordance with Dutch law. Because we are interested in children's wrong answers, seven children who gave correct answers for both of the second-order false belief questions were excluded from our analysis. Therefore, the analysis included the results of 72 children (36 female, $M_{\text{age}} = 5.7$ years, $SE = 0.05$, range: 5.0 – 6.8).¹⁴

2.3.2. Materials

Children's answers to 17 different second-order false belief stories of two different types¹⁵ were analyzed: (i) 3 'Three locations' stories, (ii) 14 'Three goals' stories. Within the story types, we always kept the structure the same while we changed the name, gender and appearance of the protagonists, along with the objects and the locations, or goals. Stories of both types were constructed in such a way that it is possible to infer whether children's possible answers to second-order false belief questions correspond to zero-order, first-order or second-order reasoning. Control questions including the reality (zero-order) and first-order false belief questions were asked before the second-order false belief questions, to test that children did not have major memory problems about the story facts, linguistic problems about the questions, and first-order false belief attribution.

'Three locations' stories were constructed based on Flobbe et al.'s (2008) 'Chocolate Bar' story (see Figure 2.1). As we discussed in the introduction, inspired by Hollebrandse et al.'s (2008), we modified Flobbe et al.'s (2008) Chocolate Bar story in such a way that it is possible to distinguish children's possible reasoning levels (i.e. zero-order, first-order, second-order) from their answers to second-order false belief questions. Before the second-order false belief question (e.g., "Where does Ayla think that Murat will look for the chocolate?") and the justification question ("Why?"), we asked four control questions. The first and second control questions were asked after Figure 2.1d as follows: i) "Does Murat know that Ayla put the chocolate into the toy box?", ii) "Does Ayla know that Murat saw her putting the chocolate into the toy box?". The third control question (zero-order ToM) was asked after the fifth episode in Figure 2.1e: iii) "Where is the chocolate now?".

¹⁴ Note that in our larger training study, children were also tested with a ToM game and a counting span task. In the ToM game, children were expected to reason about the computer's decision and about the computer's belief about their own decision. However, the task was too hard for our sample. For the counting span task, we constructed series of logistic regression models in order to test the effect of the counting span task score on children's success and failure of second-order false belief questions, and also to test its effects on the different orders of children's wrong answers (i.e. zero-order, first-order) for both of the second-order false belief questions. None of those effects were significant. For this reason, we do not present the task and its results here.

¹⁵ Children were also tested with another type of stories that are constructed based on Sullivan et al.'s (1994) Birthday Puppy story. Unlike the 'Chocolate Bar' story that we explained in the Introduction, this type of stories includes only two possible answers for the second-order false belief question. Therefore, it is not possible to distinguish whether children's wrong answers are zero-order or first-order answers. For this reason, we do not include the results of this story type here. All the tasks used in this study and the different story types were presented in random order.

Subsequently, the fourth control question (first-order false belief question) was asked: iv) "Where will Murat look for the chocolate?".

'Three goals' stories included and extended the stories used in Hollebrandse et al.'s (2008) study. One of the examples of this story type is as follows: "Ruben and Myrthe play in their room. Myrthe tells Ruben that she will go to buy chocolate-chip cookies from the bake sale at the church and she leaves the house. After that, their mother comes home and tells Ruben that she just visited the bake sale. Ruben asks his mother whether they have chocolate-chip cookies at the bake sale. The mother says, 'No, they have only apple pies'. Then Ruben says, 'Oh, then Myrthe will buy an apple pie'". At this point, the experimenter asked the first control question: "Does Myrthe know that they sell only apple pies in the market?". The story continued: "Meanwhile, Myrthe is at the bake sale and asks for the chocolate-chip cookies. The saleswoman says, 'Sorry, we only have muffins'. Myrthe buys some muffins and goes back home". Now, the second control question "Does Ruben know that Myrthe bought muffins?" and the first-order false belief question "What does Ruben think they sell in the market?" together with the justification question "Why does he think that?" were asked. Then the story proceeded: "While she is on her way home, she meets the mailman and tells him that she bought some muffins for her brother Ruben. The mailman asks her what Ruben thinks that she bought". Then, the experimenter asked the participant the second-order false belief question: "What was Myrthe's answer to the mailman?". The justification question "Why?" was asked after the second-order false belief question.

There are three possible answers to the second-order false belief question that children might report: chocolate-chip cookies, which Myrthe told Ruben initially (correct second-order answer); an apple pie, which the mother told Ruben (first-order answer); and muffins, which Myrthe really bought (zero-order answer).

2.3.3. Procedure

All the stories were presented to the children on a 15-inch MacBook Pro and were implemented with Psychopy2 v.1.78.01. All the sessions were recorded with QuickTime. If a child gave a correct answer for a second-order false belief question, his or her score was coded as 1, while incorrect answers were coded as "zero-order" or "first-order" or "I don't know", based on the given answer.

The two different types of second-order false belief stories were pseudo-randomly drawn from a pool that contained 17 different false belief stories (3 'Three location' stories and 14 'Three goals' stories). Drawings illustrating the story episodes were presented one by one, together with the corresponding

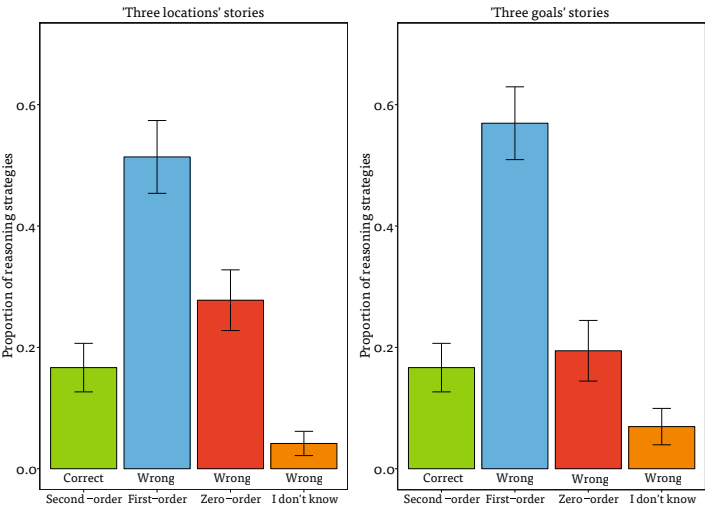


Figure 2.4. The proportion of children's level of ToM reasoning strategies when answering the second-order false belief questions (a) in 'Three locations' stories, (b) in 'Three goals' stories.

audio recordings. The drawings remained visible throughout the story. A child was never tested on the same story twice. Children did not get any feedback.

2.3.4. Results

Figure 2.4 shows the proportion of children's level of ToM reasoning for the second-order false belief questions. Confirming our instance-based learning model's prediction, most of the time children's wrong answers to the second-order false belief questions were first-order ToM answers (51% in the 'Three locations' stories, and 57% in the 'Three goals' stories) and relatively few of the answers were zero-order ToM answers (28% in the 'Three locations' stories, and 19% in the 'Three goals' stories). Overall, 17% of the second-order false belief answers were correct and 83% of them were wrong. Whereas 65% of the wrong answers were based on a first-order theory of mind strategy, 29% of them were based on a zero-order strategy, and the remaining 6% was "I don't know".

A chi-square test of independence was performed to examine the relation between the two story types and the children's levels of reasoning in their wrong answers. The relation between these variables was not significant. For this reason, we merged the data over the story types and conducted a chi-square test of goodness-of-fit to determine whether the zero-order, first-order and "I don't

know" answers were given equally often. Different levels of children's wrong answers were not equal in the population, $X^2(2, N = 119) = 64.76, p < .001$.

Table 2.3 shows the percentages of correct answers for each type of question (i.e. control, first-order false belief, second-order false belief). As can be seen from Table 2.3, children almost all the time gave correct answers for the control questions for both of the story types. Their percentage of correct answers for the first-order false belief questions was lower in 'Three locations' stories (81%) than 'Three goals' stories (93%)¹⁶. Children's correct answers to the second-order false belief questions were lower than the chance level 33% for both for the 'Three locations' stories (17%) and the 'Three goals' stories (17%).

Table 2.3. The percentages of correct answers and standard errors (in parenthesis) for the control, first-order false belief and second-order false belief questions for both 'Three locations' and 'Three goals' story types.

Questions	'Three locations'	'Three goals'
Control	95% (.02)	96% (.01)
First-order false belief	81% (.05)	93% (.03)
Second-order false belief	17% (.05)	17% (.05)

2.4. Discussion, conclusions and future work

In order to provide a procedural account for children's strategy selection while they are answering second-order false belief questions, we constructed two computational cognitive models: an instance-based model and a reinforcement learning model. Importantly, we did not introduce any additional parameters to the core cognitive architecture ACT-R to trigger a transition from incorrect to correct answers and we stated a model-based prediction before conducting our empirical study. Our main finding in this study is the confirmation of our instance-based learning model's prediction that 5- to 6-year-old children who have enough experience in first-order theory of mind but fail in second-order false belief tasks apply a first-order ToM strategy in the second-order false belief tasks. Our empirical results showed that most of the wrong answers to the second-order false belief questions were based on a first-order theory of mind strategy (65%) and few of the wrong answers were based on a zero-order strategy (29%). Note that, as we presented in Section 2.2.7.3, the reinforcement learning model did not predict that

¹⁶ The difference between the two story types were at the significance level for children's correct answers to first-order false belief questions, $X^2(1, N = 144) = 3.88, p = .05$.

children's wrong answers would most of the time be based on the first-order reasoning strategy. Before highlighting the differences between the two models and explaining their meaning in children's second-order ToM development, we would like to discuss our empirical findings compared to the previous findings in the second-order ToM literature.

Our empirical findings are consistent with Hollebrandse et al.'s (2008) results that we explained in the Introduction. Considering that children were older in Hollebrandse et al.'s study, the higher proportion of second-order ToM answers (58%) and the lack of zero-order ToM answers in their experimental results compared to ours are in line with our instance-based model's first prediction that children who have enough experience with first-order false belief reasoning but not with second-order reasoning do not give zero-order answers but first-order answers for the second-order false belief question. On the other hand, our empirical findings are in contrast with de Villiers et al.'s (2014) preliminary results showing that most five- to six-year-olds in low-income preschools gave zero-order ToM answers (60%) in the second-order false belief task, and fewer of the answers were based on the first-order ToM strategy (20%). We argue that the difference between our results and de Villiers et al.'s preliminary results can be attributed to their sample's low-income socioeconomic status, compared to our sample's upper-middle income socioeconomic status (see Cole & Mitchell, 2000; Holmes et al., 1996 for significant correlations between family socioeconomic status and individual differences in false belief performance). Moreover, the school at which we tested the children is called 'Excellence school', meaning that children's scores on the national tests are almost at the upper limit. The school's success comes from their adaptive education, which tries to ensure that both the gifted and the weaker students perform at their individual maximum. These educational and socioeconomic differences might be a possible explanation for the different results.

Our empirical findings confirm the instance-based learning model's prediction about children's wrong answers. One could argue that children may have been primed to give first-order answers because the first-order false belief questions were always asked right before the second-order false belief questions. This interpretation suggests that children only retrieved the most recent strategy chunk (i.e., first-order) or alternatively that they retrieved the most active location and gave their answers accordingly. However, even if the zero-order ToM question ("Where is the chocolate?") were asked in between the first-order and the second-order false belief questions, we believe that children would usually give first-order answers. This is because four-year-old children can already pass the first-order false belief task and as long as five-year-old children have enough experience with first-order ToM reasoning, they would usually give first-order

answers instead of zero-order answers. De Villiers et al.'s (2014) findings can be seen as evidence that asking first-order false belief questions right before the second-order false belief questions does not necessarily prime children to give first-order answers.

Moreover, the linguistic literature that shows that children respond to the embedded part of the second-order false belief question (i.e., first-order ToM reasoning) can be used as evidence that children do not only repeat the last given answer but that they have problems with selecting a strategy or with processing embedded structures (Astington et al., 2002; Villiers et al., 2014).

In addition to the empirical validation of our instance-based learning model's prediction, our modeling approach allows us to provide insight about children's development of second-order false belief reasoning. Unlike our reinforcement learning model, our instance-based learning model selects a ToM strategy which is "as simple as possible, as complex as necessary" (Meijering et al., 2014). Similar to this approach, Goodman et al. (2006) used a Bayesian Occam's razor effect to explain that children initially reason from their own perspective (zero-order ToM) in first-order false belief tasks and then with accumulated evidence revise their strategy to taking into account another agent's perspective (first-order ToM). The main difference between Goodman et al. and our modeling approach can be formulated in terms of Marr's (1982) levels of analysis. While their model takes place at the computational level only, our models also reflect the algorithmic level.

Using the "as simple as possible, as complex as necessary" approach is also in line with the previous literature on adults' ToM reasoning in strategic games in which adults are found to start applying lower levels of ToM strategies and slowly increment their level of ToM strategy when it is necessary (Camerer et al., 2004; Goodie et al., 2012; Hedden & Zhang, 2002; Wright & Leyton-Brown, 2010) and a cost-benefit approach in which a strategy with the lowest cognitive effort (cost) and the best accuracy (benefit) is selected (Payne et al., 1993; Rieskamp & Otto, 2006). Analogous to those approaches, our instance-based learning model first applies the most salient and least cognitively effortful strategy, and then if that strategy does not work increments the strategy one level higher, instead of two levels higher or more. Unlike the reinforcement learning model, this approach proposes that children's strategy revision is explicit.

One of the assumptions of our models was that initially children have a lot of experience with the zero-order ToM strategy because they perceive the world from their own perspective. On the other hand, once children have enough experience with first-order ToM reasoning they use the first-order ToM strategy in second-order false belief tasks. To validate the assumption about children's experiences with the different orders of ToM strategies, further research is needed in which children's everyday life experiences are investigated.

Although the reinforcement learning model and the instance-based learning model provided different predictions about children's systematic errors in second-order false belief tasks, both models predicted that with exposure to second-order ToM reasoning, 5-year-olds can learn to apply correct second-order ToM strategy with feedback "Wrong" without any need for further explanations of why their answer was wrong. This prediction contrasts with the first-order ToM literature showing that 4-year-olds' first-order false belief reasoning cannot be improved without further explanations (Clements et al., 2000). On the other hand, different from the reinforcement learning model, the instance-based learning model predicts that children who get feedback with explanations are more likely to revise their wrong first-order ToM strategy to the correct second-order ToM strategy. To test these model-based predictions, we are currently conducting a training study in which children hear different second-order false belief stories in two different training days with four different experimental conditions (i.e., feedback without explanation, feedback with explanation, no feedback at all, and a control condition in which children are trained with neutral stories that do not involve ToM reasoning). Confirming our instance-based learning model's predictions, our preliminary results show that children's performance from pre-test to post-test significantly increases in the feedback without explanation condition and that children who receive feedback with further explanations improve more than children who receive feedback without any further explanations (Arslan et al., 2015a). The preliminary results of our training study signal that 5-year-olds' failure in second-order false belief tasks cannot be due to maturation, related to increasing functionality of mechanisms of the brain as in Hiatt and Trafton (2015) unless there is a stimulus-triggered brain maturation.

Because our model starts to reason from its own perspective (zero-order ToM), and then takes into account another agent's beliefs (first-order ToM), and finally uses ToM recursively (second-order ToM), we can predict that children will look first to the picture that represents reality (Figure 2.1e), then to the picture that represents the first-order ToM strategy (Figure 2.1d), and finally to the picture that represents the second-order ToM strategy (Figure 2.1c). An eye-tracking study in which we can analyze children's eye movements when children are answering the second-order false belief questions can provide more insight about the underlying processes.

Does selecting the correct reasoning strategy mean that children can perfectly apply that strategy? As we explained in Section 2.2.4, once our instance-based learning model retrieves a second-order ToM strategy, it always gives the correct answer. More specifically, we have ready production rules in the model, which apply second-order ToM reasoning perfectly, when the second-order strategy has been selected. In line with the complexity explanation, we believe that selecting

the correct strategy is not the whole story in children's development of second-order false belief reasoning. When children select the correct second-order ToM strategy, they might still make mistakes for different reasons, such as lack of efficiency in applying reasoning rules and internal or external distraction.

We believe that our experimental results, which show that 29% of the wrong answers were still based on a zero-order strategy, as opposed to the 0% predicted by our instance-based learning model (but see S1 Materials showing that changing the noise value causes variation in the predicted percentages of wrong answers), is related to a working memory bottleneck or to distraction. The *serial processing bottleneck* (Verbrugge, 2009), which causes a lack of efficiency when children have to serially process embedded beliefs, might cause an inefficiency to use a second-order ToM strategy. This hypothesis suggests that children should serially process nested beliefs in a sequential manner. However, children who cannot pass second-order false belief tasks might have a lack of efficiency in serially processing embedded beliefs (for further evidence supporting the *serial processing bottleneck* in different cognitive domains, see Diamond et al., 2002; Hendriks et al., 2007; Ling et al., 2016; van Rij et al., 2010) because working memory acts as a bottleneck (Borst et al., 2010), meaning that people can only hold one chunk of information in working memory at a time. More specifically, when children try to answer second-order false belief questions, e.g., "Where does Mary think that John will look for the chocolate?", after inhibiting their own perspective, they might be holding in mind the answer of the embedded part of the question (first-order ToM reasoning), which is "Where will John look for the chocolate?". To be able to reason about Mary's false belief about John's belief, children need to use efficient rules to overcome the *serial processing bottleneck*.

How do children learn ToM reasoning strategies? We surmise two possible answers to this question. The first possible answer is related to learning common sense knowledge to reason about the false belief task. Although our model learns to pass second-order false belief tasks by repeating the task itself, in real life children do not learn second-order ToM with false belief stories. Heyes and Frith (2014) propose that explicit ToM is culturally inherited, and that parental stories and "causal-explanatory" statements might be some of the possible sources of this common sense knowledge. The second possible answer is related to learning those strategies from other cognitive tasks that are not specifically related to ToM. It is unlikely that people have complex and specialized rules in their minds to give a specific answer to false belief questions, as we have in our model. Based on Arslan et al.'s (2015b) computational cognitive modeling study which showed that working memory and cognitive control strategies can contribute to children's transitions from failure to success in first-order false belief tasks, we propose that one of the important sources of combining those complex and

specialized production rules might be children's experience in working memory strategies that they apply in their daily lives, such as counting and comparing the numbers of objects¹⁷. This explanation needs to be tested by designing a training study in which children are trained with simple and complex working memory tasks and their performance in second-order false belief reasoning from pre-test to post-test is assessed.

To sum up, unlike the reinforcement learning model, our instance-based model is able to predict 5-year-olds' systematic errors in second-order false belief tasks, namely by first-order ToM strategy. Our modeling approach provides a plausible explanation for children's systematic errors in second-order false belief tasks and shows how they revise their ToM strategy. Based on our instance-based learning model, we can surmise that 5-year-old children's failure is due to lack of experience in using a second-order ToM strategy and that children can explicitly revise their wrong first-order ToM strategy to a correct second-order ToM strategy by exposure to second-order ToM reasoning.

¹⁷ See Taatgen (2013) about the PRIMs theory for the details of how complex production rules are combined together and how experience in one domain transfers to another domain.

Chapter 3: Five-year-old Children's Development of Second-order False Belief Reasoning Can Be Accelerated: An Evaluation of Different Feedback Methods

*In which we investigate the role of different types
of feedback on children's transitions from failure
to success in second-order false belief tasks.*



**This chapter is in preparation to submit to a journal
and a short version was previously published as:**

Arslan, B., Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2015a). Teaching children to attribute second-order false beliefs: A training study with feedback. In D. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, (pp.108-113). Austin, TX: Cognitive Science Society.

Abstract

For the first time in the literature, we have conducted a training study in order to accelerate 5- to 6-year-olds' development of second-order false belief reasoning. Previous studies have shown that it is only possible to accelerate 4-year-olds' development of first-order theory of mind by providing feedback with explanations but not feedback without explanations or without any feedback. Following these findings, recent studies used feedback with explanations to accelerate 9-year olds' development of advanced theory of mind. In order to assess whether the findings of the first-order theory of mind literature still hold, we trained 106 children with second-order false belief tasks in one of the following conditions: (i) Feedback with explanation; (ii) Feedback without explanations; (iii) No feedback; (iv) Active control. The results showed that there were significant improvements in children's scores from pre-test to post-test in the three experimental conditions, from 31% to 68% in the feedback with explanation condition; from 25% to 49% in the feedback without explanation condition; and from 33% to 55% in the no feedback condition, compared to a small improvement in the active control condition (from 29% to 35%). Moreover, the improvements were not due to children's age, verbal abilities and working memory scores. Importantly, the children were able to generalize the training effect to another story type that they had not been trained on, and the training effect was stable at a follow-up session 4 months after the pre-test. Overall, our results highlight the difference between first-order and second-order theory of mind development and suggest that children can be helped over the threshold to second-order false belief reasoning by exposure to many stories and by asking them to reflect on second-order false belief questions, without providing explanations about their wrong answers.

Keywords: theory of mind; development; training study; second-order false belief reasoning; feedback

3.1. Introduction

In daily life, children are constantly in interaction with their friends and family members and children's everyday social competence is dependent on reasoning about others' mental states, such as beliefs, desires, or intentions, which can be different from their own – called *theory of mind* (ToM; Premack & Woodruff, 1978). It has been argued that being able to attribute a false belief to someone else provides evidence that a person has a theory of mind (Dennett, 1978; Wellman, 1990). Since then, false belief tasks have become some of the most applied tasks for testing children's development of ToM. A first-order false belief task examines whether children can attribute a false belief to a protagonist in a given story where the child knows the reality, while the protagonist has a false belief about the reality. Many studies have shown that, before the age of 4, most children cannot pass verbal first-order false belief tasks (Wimmer & Perner, 1983; Wellman, Cross & Watson, 2001).

Interestingly, once children are able to pass first-order false belief tasks, it takes them between one and three more years to use this false belief reasoning recursively by attributing a false belief to a protagonist who is attributing a mental state to another character in the story (Perner & Wimmer, 1985; Sullivan et al., 1994). For example, “Marieke (falsely) believes that Kevin believes that the chocolate is in the drawer”. This level of false belief reasoning is called second-order false belief reasoning. While first-order false belief reasoning is found to be related to social skills, such as deception (Sodian, Taylor, Harris, & Perner, 1992; Bosco & Gabbatore, 2017) and pretend play (Leslie, 1987), second-order false belief reasoning is found to be important in more advanced aspects of children's everyday social competence, such as idiom understanding (Caillies & Le Sourn-Bissauoui, 2013), irony understanding (Filippova & Astington, 2008; Bosco & Gabbatore, 2017), and reasoning about evidence (Astington, Pelletier, & Homer, 2002). As a concrete example, to successfully maintain a strategic lie, the liar has to reason about what the listener *knows* about what the liar *knows*, requiring second-order theory of mind (Talwar & Lee, 2008). Considering the importance of second-order ToM in children's more advanced social skills, it is important to find effective methods to accelerate children's development as well as to understand children's development of second-order ToM.

3.1.1. Training studies on first-order theory of mind: Related work

Several training studies have shown that it is possible to accelerate pre-school children's development of first-order ToM with a moderately strong effect size

(see Kloo & Perner, 2008 for a review, and Hofmann et al., 2016 for a meta-analysis). The general procedure in those successful first-order ToM training studies started by pre-testing children with first-order ToM tasks without any feedback. One or more days after the pre-test, children who could not pass all of the first-order ToM tasks at pre-test were trained either with only first-order ToM tasks (Appleton & Reddy, 1996; Clements, Rustin, & McCallum, 2000; Lecce, Bianco, Demichelli, & Cavallini, 2014a; Melot & Angeard, 2003; Slaughter & Gopnik, 1996) or with different cognitive tasks that correspond to influential factors in children's ToM development, such as language tasks (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Shulian, Yanjie, Sabbagh, & Jiaming, 2014; Zhang, Wu, & Zheng, 2008;), and executive functioning tasks (Kloo & Perner, 2003; see also Kloo & Perner, 2008 for a review). As an active control condition, children were trained either with stories that did not involve ToM reasoning or with tasks that are thought to be irrelevant for ToM reasoning. Generally, one to two weeks after the training sessions, children were post-tested to assess whether they had improved significantly since their initial performance at pre-test, compared to the control condition.

Importantly, most of the studies that trained children with ToM tasks used corrective feedback by providing further explanations during the training sessions. Unfortunately, almost none of the studies tested children again in a follow-up session a couple of months after the post-test session in order to assess whether children's improvements were stable over time (Hoffman et al., 2016). For positive exceptions, see Bianco, Lecce, & Banerjee (2016), Lecce et al. (2014a), and Lecce et al. (2014b), who tested children two months after the post-test. In our study on second-order ToM, we investigate different types of feedback to accelerate children's development of second-order false belief reasoning, and we do test children in a follow-up session four months after the pre-test session, in order to see whether their improvements are stable.

Two first-order ToM studies that trained children by providing feedback without explanations or without feedback are noteworthy for our study. Clements et al. (2000) tested 91 children between the ages 3 and 5 with first-order false belief tasks in four sessions, each 7 days apart (i.e., pre-test, training day 1, training day 2, post-test). In both the experimental and practice conditions, children heard one false belief story per day. While in the experimental condition children received feedback with detailed explanations during the training sessions, in the practice condition they only received the feedback "Correct/Wrong", without further explanation. In the control condition, children listened to neutral stories that did not involve any ToM reasoning. While children's scores in the experimental condition significantly improved from pre-test to post-test, children who were in the practice and control conditions did not show significant improvements.

Similarly, Melot and Angeard (2003) trained 93 children between the ages 4 and 5 with first-order ToM tasks. At both pre-test and post-test, children were evaluated on first-order false belief tasks and appearance-reality tasks (e.g., an imitation pencil made out of rubber). Children who did not succeed both the false belief tasks and the appearance-reality tasks were trained with only six first-order false belief tasks in one experimental condition, and with only six appearance-reality distinction tasks in a second experimental condition. In both experimental conditions, feedback together with explanations was provided during the training sessions. Children in the control group were trained with three first-order false belief tasks in one training session, and with three appearance-reality tasks in another training session. Different from the experimental conditions, children did not get any feedback in the control condition. The results showed that children's performance at the post-test had improved in comparisons to the pre-test in both of the experimental conditions but not in the control condition in which children were tested with first-order ToM tasks but did not get any feedback.

In summary, Clements et al.'s (2000) and Melot and Angeard's (2003) studies have emphasized the importance of explicit feedback with detailed explanations in 3- to 5-year-old children's improvements in first-order ToM, and have shown that children's first-order ToM performance could not be improved by providing feedback without explanations or without providing any feedback. Based on these results, as we mentioned above, almost all of the previous training studies of ToM have provided feedback with detailed explanations in the training sessions in their experimental conditions.

3.1.2. Training studies on second-order theory of mind: Related work

Given that children's ToM development goes beyond first-order false belief reasoning and continues to develop after they reach the age of 5, it is important to know whether it also holds that training children with second-order false belief tasks is only useful when feedback with explanations are provided. Moreover, understanding which types of feedback accelerate children's second-order ToM might help to understand the underlying mechanism of children's ToM development beyond the pre-school years (see Miller, 2009; 2012 for an extensive review).

However, as far as we know, there are no training studies on second-order false belief reasoning yet. There are two important previous training studies on children's ToM development beyond the pre-school years. Lecce et al. (2014b) designed a conversation-based training study for 9- to 10-year-old children to

investigate the efficacy of the conversations about mental states in children's development of advanced ToM. Because most children around the age of 9 already pass second-order false tasks, Lecce et al. used a more advanced and naturalistic ToM task – an Italian version of the Strange Stories task (Happé, 1994) – in which children's ability to make inferences about mental states in nonliteral statements was assessed. During the training sessions, children participated in a group conversation about the stories and got corrective feedback and further explanations. In the control condition, children had to reason about similar stories to the Strange Stories task, however, this time, the stories involved physical events instead of mental states reasoning. Their findings showed that children's performance from pre-test to post-test significantly improved for children in the experimental condition compared to the children in the control condition, and this improvement was stable over 2 months. Following Lecce et al.'s (2014b) findings, Bianco et al. (2016) focused on the question *if* and *how* conversations about mental states contribute 9- to 10-year-old children's development of advanced ToM. They tested two possible explanations, namely frequent use of mental-state lexicon and accuracy of mental-state attribution. They applied the same training procedure of Lecce et al.'s (2014b) study where children participated in a group discussion and got corrective feedback with further explanations. They concluded that the accuracy of mental-state attribution, but not the frequency of mental-state lexicon, mediated the positive effect of conversations about mental states on children's advanced ToM development.

In summary, both of those studies support the previous findings from the first-order ToM literature about the positive role of feedback with explanations for children's further development of ToM. However, in those two studies children did not train with a condition in which they would do advanced ToM tasks but would only get feedback "Correct/Wrong" together with the correct answer without further explanations, or would get no feedback at all. Therefore, it is still unknown whether children would still improve in those conditions. Moreover, as far as we know, there is no literature on the role of different types of feedback on second-order ToM tasks for children older than 4 but younger than 9, especially those children on the brink of developing second-order ToM.

3.1.3. The current study

In this study, we aim to fill the above-mentioned gaps in the literature by training 5- to 6-year-old children, who are on the brink of passing second-order false belief tasks, with 12 different second-order false belief stories with different types of feedback: (i) *Feedback with explanation*: by providing feedback "Correct/Wrong"

together with the correct answer and further explanations about the reason why it is the answer; (ii) *Feedback without explanations*: by providing feedback "Correct/Wrong" together with the correct answer but without further explanations; (iii) *No feedback*. To the best of our knowledge, this is both the first time that children have been tested with such a large number of different second-order false belief stories, and the first time that the roles of different types of feedback on second-order false belief reasoning have been investigated.

We have two specific hypotheses about the possible effects of training children using feedback with explanation and feedback without further explanations. Based on the previous first-order and advanced ToM training studies, we expect that children who are in the feedback with explanation condition will show an improvement in their second-order false belief scores from pre-test to post-test sessions, and that they will improve significantly more than the children who are in the control condition.

Our second hypothesis is based on a previous computational modeling study that simulated children's transitions from first-order ToM to second-order ToM (Arslan, Taatgen, & Verbrugge, 2013; 2017a). Contrary to the first-order ToM literature, one of the predictions of the computational cognitive model was that 5-year-old children who mastered first-order ToM can, in principle, pass second-order false belief tasks with the help of the feedback "Correct/Wrong" without any need for further explanation. Arslan et al. (2013; 2017a) have argued that the problem that most 5-year-old children who cannot pass second-order false belief tasks still encounter is that children are not used to reasoning about second-order mental states, therefore, they are not able to select the correct second-order ToM strategy to answer a second-order false belief question (e.g., "Where does Ayla think that Murat will look for the chocolate?"). They proposed that children can revise their wrong ToM reasoning strategy (i.e., first-order ToM) and can pass second-order false belief tasks by getting sufficient exposure to second-order false belief reasoning and getting the feedback "Correct/Wrong" without any need of further explanation. Therefore, considering the prediction of the computational modeling study, we expect that children who are in the feedback without explanation condition will also show an improvement from pre-test to post-test sessions.

In addition to second-order false belief tasks, we tested children with a working memory task (see Method section for the details). The rationale for testing children with a working memory task was based on the findings of previous literature on the role of executive functions in children's development of first-order ToM (Benson, Sabbagh, Carlson, & Zelazo, 2013; Carlson, Moses, & Breton, 2002; Carlson, Moses, & Claxton, 2004; Devine & Hughes, 2014), and second-order ToM (Perner, Kain, & Barchfeld, 2002; but see Hasselhorn, Mähler, & Grube, 2005 for no significant correlation between the working memory span score and

children's second-order false belief score when verbal abilities and age were controlled for).

Importantly, the methodology of our training study covers and extends the important suggestions of Hoffman et al.'s (2016) meta-analysis of the training studies of ToM, by: 1) controlling for working memory, verbal abilities and age; 2) testing children in a follow-up session 4 months after the pre-test session; 3) controlling for children's simple strategy use instead of reasoning about others' minds by testing them with second-order true belief stories; 4) testing children with a ToM task of a type that is not used in the training sessions in order to see whether children can generalize what they have learned; and 5) using an active control condition.

3.2. Method

3.2.1. Participants

One hundred and nineteen children were recruited from a primary school with predominantly upper-middle-class families from Groningen, the Netherlands. All children had Dutch as their first language, and they were students of three different teachers. All different conditions contained about equal numbers from all these three teachers who teach 5- to 6-year-olds. The analysis showed that there was no significant effect of the teacher in children's scores at pre-test and improvements after the training sessions, and adding teachers as a random effect did not improve the linear mixed effect models. Therefore, we merged the data across the teachers for the rest of the analysis. We sent a written parental consent to the parents via the teachers. The children whose parents did not object to participation in the experiment and who did not have cognitive or learning difficulties were initially included. Children were pre-tested to ensure that they had not yet fully developed second-order false belief reasoning. Thirteen children were excluded from the study, as follows. Nine of children (aged 5;0, 5;3, 5;3, 5;4, 5;5, 5;8, 5;8, 5;8, 6;1) were already good at second-order false belief reasoning and gave correct answers for all the three second-order false belief questions. Two children (aged 5;4, 5;8) left the study before it was completed; moreover, one child was excluded due to technical problems during the experiment (aged 5;5), and one child (aged 5;1) was excluded because she was not able to answer any of the first-order false belief questions at the pre-test. Thus, the analysis included the results of 106 children in three experimental conditions and one control condition. Table 3.1 illustrates the number of participants (the number of female participants in parentheses), the age range, and the mean age (standard deviations in parentheses), the verbal

ability scores (standard errors in parentheses), the working memory score at pre-test (standard errors in parentheses) in each condition.

Table 3.1. The number of participants (the number of female participants in parentheses), the age range, and the mean age (standard deviations in parentheses), the verbal ability scores (standard errors in parentheses), the working memory score at pre-test (standard errors in parentheses) in each condition

Condition	N	Age range	M _{age} (SD)	Verbal ability (SE)	WM _{pre-test} (SE)
Feedback with explanation	23 (15 female)	5;1 – 6;2	5;8 (0.28)	51.02 (0.10)	2.45 (0.21)
Feedback without explanation	23 (10 female)	5;2 – 6;8	5;8 (0.44)	53.70 (0.08)	1.98 (0.25)
No Feedback	26 (11 female)	5;2 – 6;8	5;4 (0.25)	53.01 (0.08)	2.25 (0.22)
Control	34 (19 female)	4;8 – 6;5	5;3 (0.35)	51.51 (0.13)	2.17 (0.21)

3.2.2. Design

Children were tested in three different experimental conditions and one control condition: (i) Feedback with explanation; (ii) Feedback without explanation; (iii) No feedback; (iv) Control. Each child was tested in five separate sessions, namely pre-test, training day 1, training day 2, post-test, and follow-up. There was at least one day intermission between the pre-test, training day 1, training day 2, and post-test sessions, and there was at least one week and at most nine days of intermission between the pre-test and the post-test sessions. The follow-up session was conducted four months after the pre-test session. Fig. 3.1 shows the design of the experiment.

At the pre-test, post-test and follow-up sessions, children were tested with a working memory task and three second-order false belief stories (1 'Three goals', 1 'Decoy gift', and 1 'Three locations') in a random order¹⁸. In the three different experimental conditions, children heard 6 second-order false belief stories (3 'Three goals', 3 'Decoy Gift') at each training session. After the third and the sixth second-order false belief stories, they heard one second-order true belief story (1 'Three goals', 1 'Decoy gift'). In each second-order false belief story and

¹⁸ Children were also tested with a theory of mind game at pre-test, post-test and follow-up sessions in order to investigate whether they could generalize what they had learned and apply second-order ToM in a very different context. In this game, children were expected to reason about the computer's decision (first-order ToM) and about the computer's decision about their own decision (second-order ToM) to gain the highest possible outcome. However, the task was too hard for the 5-6 year olds. For this reason, we do not present the game and its results here.

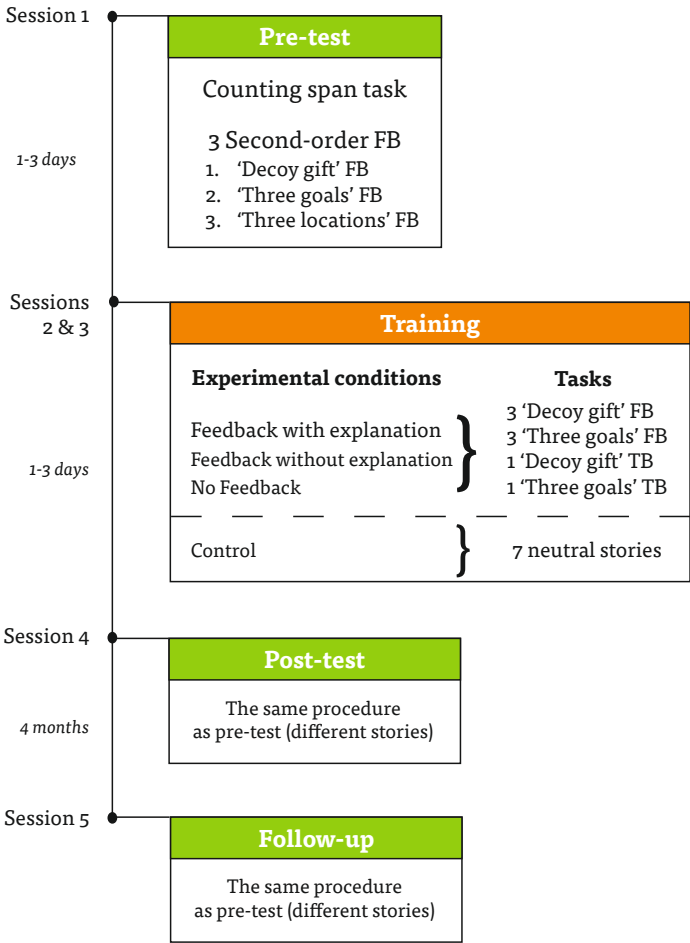


Figure 3.1. The design of our training study.

each second-order true belief story of a certain type, we fixed the general story structure, but we changed the protagonists' gender, appearance and name, as well as objects, locations and further context of the stories. In the control condition, in both training sessions, children were tested with 7 neutral stories that do not involve any level of false belief reasoning. Each neutral story had approximately the same length as the second-order false belief stories and the second-order true belief stories.

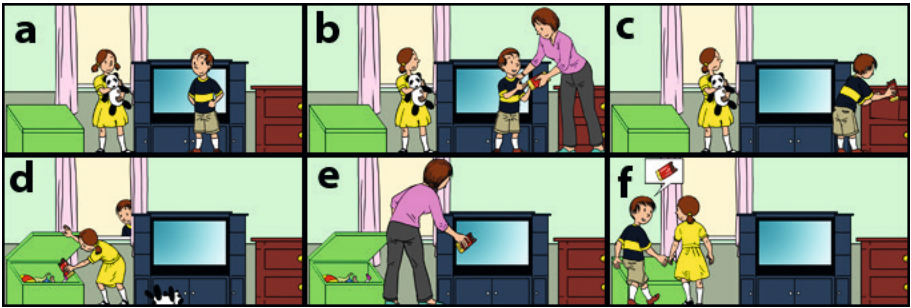


Figure 3.2. The prototype 'Three locations' story, namely the 'Chocolate Bar' story (Illustration ©Avik Kumar Maitra)

a) Kevin and Marieke are brother and sister. They are in the living room.
b) Their mother bought a chocolate bar and gives it to Kevin. Marieke doesn't get any chocolate, because she has been naughty.
c) Kevin eats some of his chocolate and puts the remainder into the drawer. He doesn't give any chocolate to Marieke. Marieke is upset that she does not get any chocolate.
d) After that, Kevin goes to help his mother in the kitchen. Marieke is alone in the room. Because she is upset, she takes the chocolate from the drawer, and puts it into the toy box. While she is putting the chocolate into the toy box, Kevin is passing by the window. He sees how Marieke takes the chocolate out of the drawer and puts it into the toy box. Marieke does not see Kevin.

At this point, the pre-recorded control questions "Does Kevin know that Marieke put the chocolate into the toy box?", and "Does Marieke know that Kevin saw her put the chocolate into the toy box?" were asked.

e) After that, Kevin goes back to the kitchen and Marieke goes to the kitchen, as well. While Kevin and Marieke are in the kitchen, their mother goes to the living room to watch TV. While she is searching for the remote control, she sees the chocolate in the toy box. The mother is surprised that the chocolate is in the toy box. She takes the chocolate from the toy box and puts it into the TV stand. She watches TV for a while and goes to her room.

At this point of the story, the reality control question "Where is the chocolate now?" was asked.

f) Now, Kevin and Marieke go back to the living room. Kevin wants to eat some of his chocolate. He says: 'Hmm, I would like to some chocolate'.

At this point the first-order false belief question "Where will Kevin look for the chocolate?" and the justification question "Why does he look there?" were asked.

Subsequently, the second-order false belief question: "Where does Marieke think that Kevin will look for the chocolate?" was asked together with the justification question "Why does she think that?"

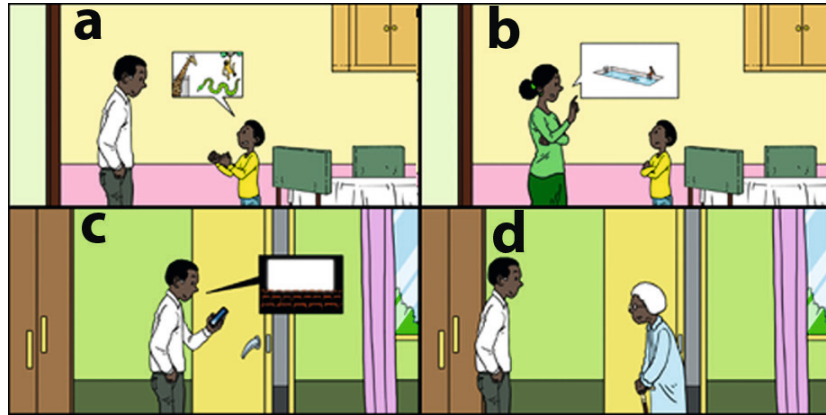


Figure 3.3. The drawings of the prototype of 'Three goals' stories, namely 'A Day Out' story (Illustration ©Avik Kumar Maitra)

a) It is Robert's birthday, so Robert's dad promised to do something fun. Dad asks 'Where do you want to go today?'. Robert says: 'The zoo!'. Dad wants to call the zoo in order to make sure that it is open. He walks out of the room to get his phone.

b) Then, mother comes to the room. She asks Robert: "What are you doing today?". Robert says "We will go to the zoo!". Mom says: "The zoo is not open today but you can also go to the swimming pool". Robert thinks this is a good idea. He goes to find his dad to tell him that he wants to go to the swimming pool.

At this point, the control question "Does dad know that Robert wants to go to a swimming pool?" was asked.

c) Dad is alone in his room and he calls the zoo. He learns that the zoo is closed today. What now? He says to himself: "I know where to go, there is a very good movie in the cinema today, so I will call and book tickets for the movie".

At this point the second control question "Does Robert know that his dad will go to a movie with him?" and the first-order false belief question "What does Robert think they are going to do today with his dad?" together with the justification question "Why does he think that?" were asked.

d) When dad has reserved the movie tickets, grandmother comes inside. She asks "What will you do with Robert today?". Dad says: "We will go to the cinema". Grandma says: "Oh, does Robert know what you are going to do today?".

At this point, the control question (ignorance) "What does dad say to grandma?" was asked.

Subsequently, the last part of the story was told: "Then the grandma asks: "What does Robert think that you will do today?".

At this point, the second-order false belief question "What does dad say to grandma?" together with the justification question "Why does she say that?" were asked.

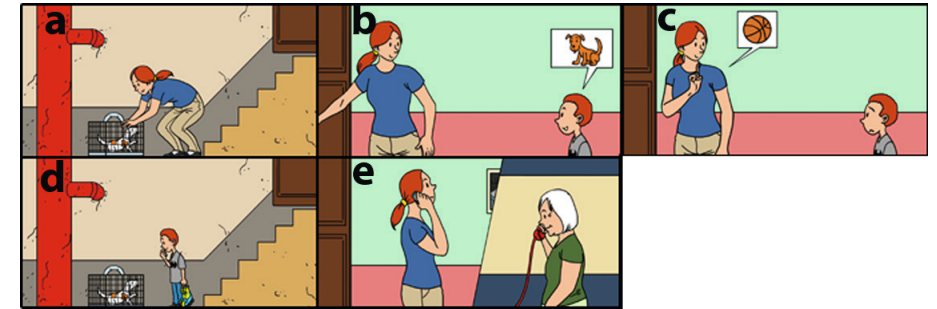


Figure 3.4. The prototype example of 'Decoy gift' stories, namely the 'Birthday Puppy' story (Illustration ©Avik Kumar Maitra)

a) Tonight, it's Rick's birthday and his mum wants to surprise him with a puppy. She has hidden the puppy in the basement.

b) Rick says, "Mum, I really hope you got me a puppy for my birthday".

c) Because Rick's mother wants to surprise him with a puppy, instead of telling Rick she got him a puppy, she says: "Sorry Rick, I didn't get you a puppy for your birthday. I got you a really nice basketball instead".

At this point, the control question "What did the mother really get Rick for his birthday" and the first-order false belief question "What does Rick think that his mom bought for him?" together with the justification question "Why does Rick think that?" were asked.

d) Now, Rick says to his mother: "I am going outside to play". On his way outside, Rick goes down to the basement to fetch his skates. In the basement, Rick finds his birthday puppy. Rick says to himself: "Wow, mum didn't get me a basketball; she really got me a puppy for my birthday". His mother does not see that Rick goes down to the basement and finds the birthday puppy.

At this point the control question "Does Rick know that his mother got him a puppy for his birthday?" was asked.

e) Now the telephone rings, ding-a-ling! Rick's grandmother calls to find out what time the birthday party is. The mother tells grandma on the phone that she got Rick a puppy for his birthday, but that Rick doesn't know this. Then, grandma asks mum on the phone, "What does Rick think you got him for his birthday?".

Subsequently, the second-order false belief question "What does the mother say to grandma?" together with the justification question "Why does mum say that?" were asked.

3.2.3. Materials

3.2.3.1. Second-order false belief stories

We constructed 31 different second-order false belief stories of three different types: (i) 3 'Three locations' stories, (ii) 14 'Three goals' stories, (iii) 14 'Decoy-gift' stories. For all stories, children were asked a question that required second-order false belief attribution, as well as some control questions. In the literature, second-order false belief questions often have two possible answers, for example, two locations. We constructed 'Three locations' and 'Three goals' stories in such a way that our second-order false belief questions have three different possible answers, according to which we can distinguish children's level of reasoning (i.e., zero-order, first-order, second-order).

Figure 3.2 shows the prototype example of 'Three locations' stories, namely the 'Chocolate Bar' story. These 'Three locations' stories were constructed based on Flobbe, Verbrugge, Hendriks, and Krämer's (2008) version of Hale and Tager-Flusberg's (2003) 'Chocolate Bar' story. There are three possible answers to be reported to the second-order false beliefs question "Where does Marieke think that Kevin will look for the chocolate?": i) second-order ToM answer: "the drawer", because Marieke *thinks that* Kevin *thinks that* the chocolate is in the drawer; ii) first-order ToM answer: "the toy box", because Kevin *thinks that* the chocolate is in the toy box; iii) zero-order ToM answer: "the TV stand", because the chocolate is in the TV stand.

Figure 3.3 shows the prototype example of 'Three goals' stories, namely 'A Day Out'. 'Three goals' stories included and extended the stories used in Hollebrandse, van Hout, and Hendriks' (2014) study. Just like in the 'Three location' stories, there are three possible answers to the second-order false belief question: i) second-order ToM answer: "the zoo", because Dad *thinks that* Robert *thinks that* they will go to the zoo; ii) first-order ToM answer: "the swimming pool", because Robert *thinks that* they will go to the swimming pool; iii) zero-order ToM answer: "the cinema", which is the real place to which they will go.

Figure 3.4 shows the prototype example of 'Decoy gift' stories, namely the 'Birthday Puppy' story. 'Decoy gift' stories were constructed based on Sullivan et al.'s (1994) Birthday Puppy story. Unlike the 'Three locations' and 'Three goals' stories, in this story, there are two answers that the participants might report: i) second-order ToM answer: "a basketball", because Mother *thinks that* Rick *thinks that* she bought a basketball; ii) zero-order ToM and first-order ToM answer: "a puppy", because it is the real present and because Rick *thinks that* his mother bought a puppy.

For each story, a judgment score of 1 was given for a correct answer to a second-order false belief question, and a score of 0 was given for a wrong answer.

Similarly, if a child's justification answer included the correct information that one character does or does not know about the other character's history of exposure to relevant information, it was coded as correct (1 points). Otherwise, the justification was coded as incorrect (0 points).

3.2.3.2. Second-order true belief stories

Second-order true belief stories were used only in the training sessions. In them, children were asked to answer a question that required attribution of a second-order true belief. Because we only trained children with 'Decoy gift' and 'Three goals' stories, we constructed the second-order true belief versions for only these types of stories, namely 2 'Decoy gift' true belief stories and 2 'Three goals' true belief stories. The true belief stories have the same structure as the false belief stories. However, the protagonist whose belief the child has to report entertains a true belief instead of a false belief. For instance, in the true belief story corresponding to the 'Decoy Gift' story given above, the son finds his real birthday present, but the mother is also in the room and they jointly attend the present. Therefore, this time the correct answer (a puppy) to the second-order true belief question is not the same as the correct answer to the second-order false belief question in the corresponding false belief story (a basketball), because now the mother knows that the son knows that what she bought for him. For each story, a judgment score of 1 was given for a correct answer to a second-order true belief question, and a score of 0 was given for a wrong answer.

3.2.3.3. Neutral stories

Neutral stories were presented to participants in the control condition in two training sessions (i.e., 7 stories in each training day). 14 neutral stories that have a similar length with the second-order false belief stories, and that do not involve theory of mind reasoning were selected from a children's book called 'Jip en Janneke' by Annie-Schmidt, with Fiep Westendorp's (2011) illustrations. Each story was divided into two episodes and presented on the computer with two drawings from the book illustrating the episodes. After each episode, two neutral questions not involving any mental state expressions were asked related to the episode of the story, in order to check if the children paid attention.

3.2.3.4. Working memory task

As a working memory task, we chose to use a task that involves minimal language. For this reason, we used a computerized version of the counting span task at pre-test¹⁹. We adapted the task from Towse et al.'s (1998) study. In this task, cards that have red triangles and blue squares were shown on the computer screen one by one. Children were instructed to count aloud the blue squares by pointing at them and to remember their total number on each card. The experimenter told them that after they counted the targets on the first card, the next card would be shown on the screen and they should repeat the same procedure, remembering the number of blue squares on both cards. After being sure that children understood the instructions and practiced one two-cards trial, which was shown on paper with the help of the experimenter, the real experiment was shown on the computer.

In the first level, after two cards, the children were asked to report the total number of blue squares per card in the same order that the cards had been presented. Each level had three trials. If a child reported all numbers back correctly for a trial, positive feedback was provided in the form of an audio file saying "Well done!" together with a green happy smiley on the screen. If a child was not able to report all the target numbers correctly, a neutral face together with an audio "Let's try another one!" was presented. If a child correctly reported two out of three trials at a given level, then the difficulty was increased to a higher level, meaning that the number of cards per trial was increased by one. For the scoring, we adopted the criteria of Towse et al.'s (1998) study. In this scoring procedure: (i) the highest level (number of cards) for which at least two of the three trials were correct was noted as the main part of the score; (ii) if one of the three trials at the next level was correct, because this represents "half-way" toward the next span level, an additional 0.5 marks were given; (iii) the number of correct answers (correct item in the correct serial position) in each remaining trial was divided by the number of items to be recalled. The mean proportion of correct recalls for the incorrect trials was then derived. This value was multiplied by 0.5 and the product added to the total score obtained from procedures (i) and (ii).

3.2.3.5. Verbal abilities

Children's verbal ability scores were taken from their school's database and used as control variables in our statistical analyses to assess children's improvements

¹⁹ Because we wanted to keep the structure of the sessions the same, we also tested children with the counting span task at post-test and follow-up sessions. For the purpose of this study, we only include children's pre-test scores in the analyses as a fixed factor and do not report the results of the counting span score at post-test and follow-up sessions.

after the training sessions. These scores are part of the monitoring system, called CITO, for schools in the Netherlands. Starting from Grade 1 to Grade 8 (4 – 12 year-old), most of the children in the Netherlands are tested with the same instruments in order to assess children's progress systematically. Children's verbal abilities were tested in terms of vocabulary and answering a question after listening to a small story.

3.2.4. Procedure

Children were tested individually in their school in a separate room by one of seven experimenters. Because children had not yet learned how to read, all the stories and the questions were presented via the computer's speakers. All the drawings and the audios files were implemented in Psychopy2 v.1.78.01 and were presented to the children on a 15-inch MacBook Pro OS X 10.10.5. All experimenters were trained before running the experiment in order to follow the same instructions. In the results, it turned out that there were no significant differences in scores between the children who were trained with different experimenters and adding experimenters as a random effect did not improve the linear mixed effect models. Therefore, we merged the data for the rest of the analysis. A child was almost always tested by the same experimenters at the pre-test, training day 1, training day 2, post-test and follow-up sessions. As an exception, two of the experimenters were not able to attend the follow-up test for 17 children, which was four months after the post-test. Thus, two of the remaining five experimenters tested these 17 of 106 children for the follow-up session. Each session took approximately 30 minutes. All of the sessions were recorded with QuickTime's screen recording together with the audio recording. After each session, children received three stickers for "doing so well."

The stories were drawn randomly, without repetitions from a pool that contained 31 different second-order false belief stories, a pool of 4 different second-order true belief stories, and (for the control condition) a pool of 14 different neutral stories. Drawings illustrating the story episodes of the stories were presented one by one, together with the corresponding audio recordings. The drawings remained visible when children were asked questions. As is usual in the previous studies, control questions were asked before the second-order false belief and second-order true belief questions in order to test that children did not have major memory and linguistic problems about the stories and the structure of the questions (Sullivan et al., 1994; Wimmer & Perner, 1983). Also, first-order false belief questions were asked before the second-order false belief questions, in order to make sure that the children did not have any major problems with first-order

false belief reasoning. If a child gave a wrong answer for a control question or a first-order false belief question in the second-order false belief tasks and the questions in the neutral stories that were used in the control condition, those questions were asked up to three times altogether by repeating the related story episodes. If a child still gave a wrong answer, it was coded as wrong. A child was never tested on the same story twice.

3.2.4.1. Pre-test, post-test and follow-up testing sessions

Children were tested with a counting span task and 3 second-order false belief stories (1 'Three goals' 1 'Decoy gift', and 1 'Three locations') in a random order. The presentation of the order of the tasks and the order of the story types were randomized. Children did not get any feedback in the pre-test, post-test, and follow-up sessions. Because children were not trained with 'Three locations' stories at the two training sessions, this type of stories was used to test whether children can generalize what they learned in the training sessions to another type of second-order false belief task.

3.2.4.2. Training Sessions

In the second and the third sessions (training day 1, training day 2), children in three of the experimental conditions were trained using six different second-order false belief stories (3 'Three goals', and 3 'Decoy gift') per training session. In addition to the second-order false belief stories in training sessions, children were tested with two second-order true belief stories in order to capture whether a child could have applied a simple strategy instead of reasoning about the second-order false belief questions. Each true belief story was presented after the 3 second-order false belief stories.

In the feedback with explanation condition, the feedback "Correct/Wrong" together with an explanation was provided in an interactive fashion. For example, the explanation was in the following form²⁰ for the prototype of the 'Decoy gift' stories that was given in the Materials section: "Correct/Wrong, Rick saw the puppy in the basement but his mom didn't see Rick, right? That's why mom tells grandma that Rick thinks he will have a basketball. Because that is what the mom told Rick before, right?". Similarly, the explanation was in the following

²⁰ Although we trained the experimenters to give the exact same feedback for each child and we provided them a script, because the feedback was provided in an interactive fashion, there were small variations in the form of the feedback between and within the participants. However, very similar information was given to each child even in case of such small deviations.

form for the prototype of the 'Three goals' stories that was explained in the Materials section: "Robert told his dad that he wanted to go to the zoo. Then, mom told Robert that the zoo is not open today and they can go to the swimming pool but dad did not hear that, right? That is why, dad says to grandmother that Robert thinks they are going to the zoo, right?"

In the feedback without explanation condition, only the feedback "Correct/Wrong" was provided, together with the correct answer without any further explanation. In the no feedback condition and in the control condition, children did not get any feedback.

3.3. Results

Figure 3.5 shows (a) proportion of correct answers to the second-order false belief questions at pre-test, post-test and follow-up sessions and (b) the difference in the proportions between pre-test and post-test sessions for each condition. There is a considerable improvement of children's scores from pre-test to post-test in the three experimental conditions: from 31% to 68% in the feedback with explanation condition; from 25% to 49% correct in the feedback without explanation condition; and from 33% to 55% in the no feedback condition, compared to a

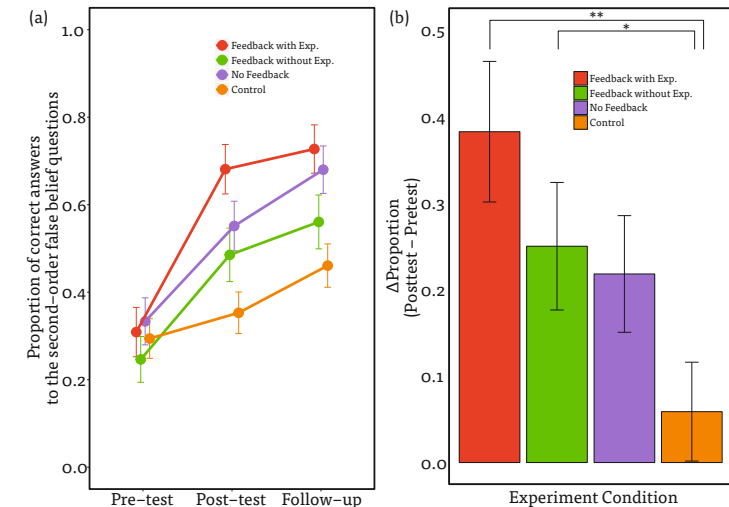


Figure 3.5. a) Proportion of correct answers to the second-order false belief questions at pre-test, post-test and follow-up sessions. b) the difference in the proportions between pre-test and post-test sessions for each condition. * $p < .05$, ** $p < .008$.

Table 3.2. The percentage (and number in parentheses) of children showing an improvement, no change or deterioration in (a) answers to second-order false belief question and (b) answers to justification questions from pre-test to post-test sessions

	Improvement		Stability		Deterioration	
	(a)	(b)	(a)	(b)	(a)	(b)
Feedback with explanation (N=23)	74 (17)	61 (14)	9 (2)	39 (9)	17 (4)	0
Feedback without explanation (N=23)	57 (13)	57 (13)	35 (8)	35 (8)	8 (2)	8 (2)
No feedback (N=26)	58 (15)	50 (13)	30 (8)	42 (11)	12 (3)	8 (2)
Control (N=34)	32 (11)	26 (9)	50 (17)	65 (22)	18 (6)	9 (3)

small improvement in the control condition (from 29% to 35%). Moreover, children who were in the experimental conditions performed better than the children who were in the control condition in the follow-up session, which was after 4 months from the pre-test session (73% correct in the feedback with explanation condition; 56% in the feedback without explanation condition; 68% in the no feedback condition; compared to 46% in the control condition).

In this section, we first present the results of the training effects from pre-test to post-test in Subsection 3.3.1. Then we investigate the generalizability of the training effects by focusing on children's improvements from pre-test to post-test sessions in Subsection 3.3.2. Subsequently, in Subsection 3.3.3, we present the stability of the training effects by focusing on the improvements from pre-test to follow-up sessions. Finally, in Subsection 3.3.4, we present the results of non-experimental questions.

3.3.1. Training effects from pre-test to post-test sessions

Following a similar pattern as Figure 3.5, Table 3.2 shows the percentage (and number in parentheses) of children showing an improvement, no change or deterioration in (a) answers to second-order false belief question and (b) answers to justification questions from pre-test to post-test sessions. As can be seen from Table 3.2, the most improvement in children's answers to second-order false belief questions and justifications answers (together with less stability and less deterioration) occurs in the feedback with explanation group. In the feedback without explanation and no feedback conditions, children showed similar patterns of improvements. Moreover, as we expected, in the control condition, children's improvement was much less, whilst their stability and deterioration were more compared to the children who were in one of the three experimental conditions.

Table 3.3. The estimates and z-values of the binomial mixed-effects model from pre-test to post-tests sessions*

	B	SE	z	p
1 (Intercept)	-1.32	0.31	-4.24	< .001
2 Post-test	0.08	0.43	0.18	.86
3 Feedback without explanation	-0.67	0.46	-1.46	.14
4 Feedback with explanation	-0.35	0.42	-0.83	.41
5 No Feedback	0.18	0.37	0.48	.63
6 'Decoy gift'	1.63	0.25	6.51	< .001
7 'Three locations'	-0.22	0.35	-0.64	.52
8 Age	0.94	0.39	2.43	.02
9 Verbal ability	-0.01	0.02	-0.60	.55
10 Working memory	0.08	0.11	0.72	.47
11 Post-test x Feedback without explanation	1.51	0.70	2.17	.03
12 Post-test x Feedback with explanation	1.88	0.70	2.71	.007
13 Post-test x No Feedback	1.07	0.64	1.68	.09
14 Control x 'Three locations' x Post-test	0.34	0.63	0.54	.59
15 Feedback without explanation x 'Three locations' x Post-test	-0.86	0.78	-1.11	.27
16 Feedback with explanation x 'Three locations' x Post-test	0.28	0.72	0.39	.70
17 No Feedback x 'Three locations' x Post-test	-0.06	0.66	-0.09	.92

* The pre-test session, 'Three goals' stories, and control condition were used as base levels in the model (reference categories).

A binominal mixed effects model was fitted on the scores with the following effects: the main effects of and interaction between session (pre-test/post-test) and condition (feedback with explanation, feedback without explanation, no feedback, control) to test for differential learning effects of the different training regiments; a three-way interaction between condition, 'Three locations' items and session to test whether learning on new items was different from old items; the centered age of the child, the centered scores for verbal ability and the centered scores for working memory capacity. As random effects, we had random slopes for session per subject correlated with the random intercepts. Table 3.3 presents the estimates of the coefficients (reported in log odds) and z-statistics of the model. Note that 'Three location' stories were not used in the training sessions. The pre-test session, 'Three goals' stories, and control condition were used as base levels in the model (reference categories).

As can be seen from Table 3.3, children's scores did not significantly improve from pre-test to post-test in the control condition (row 2). Children's second-order

false belief scores significantly improved in the feedback with explanation condition and in the feedback without explanation condition compared to children's improvements in the control condition (rows 11 and 12). There was a marginally significant improvement from pre-test to post-test sessions in the no feedback condition compared to the control condition (row 13).

As expected, there was also a significant effect of age (row 8). We did not find a significant effect of children's verbal abilities (row 9) and working memory score at pre-test (row 10) on children's second-order false belief score. We interpret the results about the different story types in the following subsection about generalizability.

3.3.2. Generalizability of the training effect

In order to investigate the generalizability of the training effect, we focus on children's improvements from pre-test to post-test sessions in 'Three locations' stories. Note that, unlike 'Three goals' and 'Decoy gift' stories, we did not test children with 'Three locations' stories at the training sessions.

As can be seen from Table 3.3 (rows 6 and 7), while there was no significant difference between 'Three goals' and 'Three locations' stories (row 7), children's scores in 'Decoy gift' stories were significantly better than children's scores in 'Three goals' stories (row 6). Moreover, although children's improvement in 'Three locations' stories in the feedback without explanation condition was not as great as the improvement in the other conditions, there were no significant differences between children's scores in the trained stories and children's 'Three locations' post-test scores in all of the conditions.

In Figure 3.6, we merged the stories that we used at the training sessions ('Other Stories'), namely 'Three goals' and 'Decoy gift' stories, and compared with 'Three locations' stories. As can be seen from Figure 3.6, for the experimental conditions, both the 'Three locations' stories and the 'Other Stories' have a similar amount of increase in the proportion of correct second-order false belief answers from pre-test to post-test (a rise of 29 percentage points in 'Three locations', and a rise of 27 percentage points in 'Other Stories' in the experimental conditions; compared to a rise of 8 percentage points in 'Three locations' and a rise of 4 percentage points in 'Other Stories' in the control condition). In more detail, Figure 3.7 shows children's improvements in all of the story types of second-order false belief stories from pre-test to post-test sessions in all conditions. These results show that children were able to generalize what they learned at the training sessions to another story type that they did not train with, namely 'Three locations'.

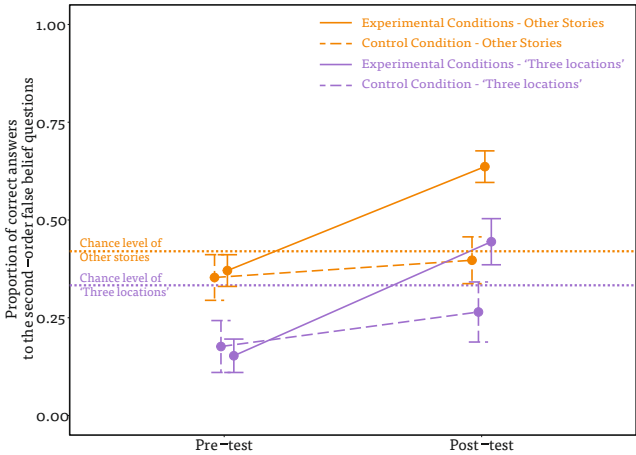


Figure 3.6. The comparison of children's improvements in 'Three locations' vs. 'Other Stories' story types of second-order false belief stories from pre-test to post-test sessions.

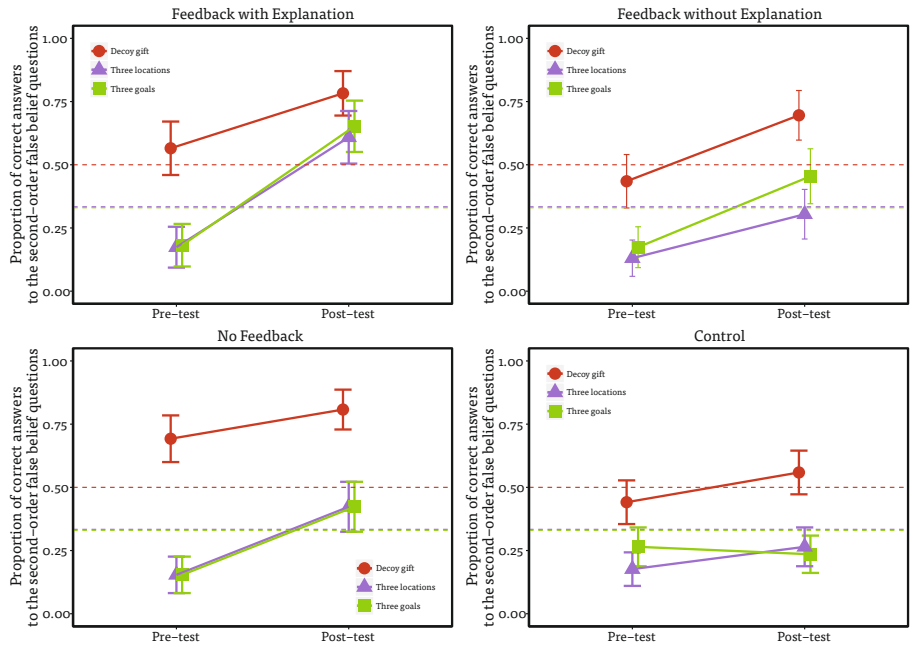


Figure 3.7. Children's improvements in all of the story types of second-order false belief stories from pre-test to post-test sessions in all conditions. The green dashed horizontal line represents the chance level for the 'Decoy gift' stories (50%), and the blue and red dashed lines represent the chance level for the 'Three goals' and 'Three locations' stories (33%).

Table 3.4. The estimates and z-values of the binomial mixed-effects model for the stability of the training effect.

		B	SE	z	p
1	(Intercept)	-1.32	0.31	-4.26	<.001
2	Follow-up	0.35	0.39	0.92	.36
3	Feedback without explanation	-0.71	0.45	-1.58	.11
4	Feedback with explanation	-0.34	0.42	-0.81	.42
5	No Feedback	0.16	0.38	0.41	.68
6	'Decoy gift'	1.64	0.24	6.75	<.001
7	'Three locations'	-0.22	0.35	-0.64	.53
8	Age	0.96	0.34	2.78	.005
9	Verbal ability	0.002	0.02	0.10	.92
10	Working memory	0.06	0.10	0.58	.56
11	Follow-up x Feedback without explanation	0.76	0.62	1.22	.22
12	Follow-up x Feedback with explanation	1.47	0.62	2.38	.02
13	Follow-up x No Feedback	0.85	0.58	1.48	.14
14	Control x 'Three locations' x Follow-up	1.38	0.57	2.41	.02
15	Feedbackwithoutexplanationx'Threelocations'xFollow-up	1.39	0.70	2.01	.04
16	Feedback with explanation x 'Three locations' x Follow-up	0.92	0.71	1.30	.19
17	No Feedback x 'Three locations' x Follow-up	1.62	0.68	2.40	.02

3.3.3. Stability of the training effects: Improvements from pre-test to follow-up sessions

As can be seen from Figure 3.5, for all the conditions, children's scores on second-order false belief stories improved from the pre-test to a follow-up session, which was 4 months after the pre-test session (a rise of 42 percentage points in the feedback with explanation condition; a rise of 31 percentage points in the feedback without explanation condition; a rise of 35 percentage points in the no feedback condition; compared to a rise of 17 percentage points in the control condition).

Similar to the fitted binomial linear mixed effect model that we presented in Table 3.3 (Model 2), in order to test the stability of the training effect, we fitted a binominal mixed effects model on the scores with an interaction between session (pre-test/follow-up) and condition (feedback with explanation, feedback without explanation, no feedback, control); an interaction between condition and 'Three locations' scores at follow-up session; and story types ('Three locations', 'Three goals', 'Decoy gift'), centered age, centered pre-test working memory scores, and

Table 3.5. The percentages of correct answers (standard errors in parentheses) for the non-experimental questions

Question type	Total number of questions	Correct answers
Control questions	4968	98% (0.002)
First-order false belief	1811	95% (0.005)
First-order true belief	286	97% (0.01)
Second-order true belief	286	92% (0.02)
Questions in neutral stories in the control condition	1900	95% (.005)

centered verbal ability score as fixed factors. As random effects, we had random slopes for session per subject correlated with the random intercepts. Table 3.4 shows the estimates and z-values of the binomial mixed-effects model for the stability of the training effect.

As can be seen from Table 3.4, in control condition, children's second-order false belief scores did not significantly improve from pre-test to follow-up sessions (row 2). There was a significant difference from pre-test to follow-up sessions between the control condition and the feedback with explanation (row 12). Children in the feedback without explanation condition and the no feedback condition performed better than children in the control condition at the follow-up session, however the differences in improvements between those conditions and the control condition were not significant (row 11, 13).

Moreover, children's improvements from pre-test to follow-up sessions in 'Three locations' stories were more than children's improvements in the trained stories in all the conditions (rows 15,16,17).

Finally, similar to the findings from pre-test to post-test sessions that is shown in Table 3.3, while age had a significant effect, children's verbal abilities and working memory score at pre-test did not have a significant effect on children's second-order false belief score from pre-test to follow-up sessions.

3.3.4. Non-experimental questions

Table 3.5 shows the percentages of correct answers (standard errors in parentheses) for the non-experimental questions. Control questions and first-order false belief questions were asked before the second-order false belief questions in order to make sure that children did not have problems with memorizing the story facts and understanding the relatively complex second-order false belief

questions as well as attributing first-order false beliefs. Moreover, in order to make sure that children did not use a simple strategy instead of reasoning about the second-order false belief questions, we investigated children's performance on second-order true belief questions.

As can be seen from Table 3.5, children predominantly gave correct answers to these non-experimental questions, meaning that they did not have problems with memorizing the story facts, nor with first-order false belief reasoning. The high proportion of second-order true belief answers shows that children did not use simple strategy instead of applying second-order ToM reasoning.

3.4. Discussion and future research

For the first time in the literature, the roles of different types of feedback, namely feedback with explanations, feedback without explanation, and no feedback have been studied, to accelerate 5- to 6-year-olds' development of second-order false belief reasoning. Crucially, the design of our study covers and extends the important suggestions of Hoffman et al.'s (2016) meta-analysis of the training studies of ToM, by: 1) controlling for age, working memory and verbal abilities; 2) testing children in a follow-up session after 4 months from the post-test session; 3) controlling for children's simple strategy use instead of reasoning about others' minds by testing them with second-order true belief stories; 4) testing children with a ToM task that is not used in the training sessions in order to see whether children can generalize what they have learned; and 5) using an active control condition.

Overall, our results draw attention to the similarities and the differences in first-order and second-order ToM development. We first discuss the training effect and its stability. Subsequently, we discuss the results about the generalizability of the training effect.

3.4.1. The training effects and their stability

As we predicted, children in the feedback with explanation condition made greater gains in second-order ToM from pre-test to post-test than children in the active control group did, when the effects of age, verbal abilities, and working memory are controlled for. The positive training effect of feedback with explanation is in line with the previous findings of first-order ToM training studies (Appleton & Reddy, 1996; Lecce et al., 2014a; Clements et al., 2000; Melot & Angeard, 2003) and with the second-order ToM studies that tested

9- to 10-year-olds with more advanced ToM tasks (Bianco et al., 2015; Lecce et al., 2014b). This result of 5- to 6-year olds shows the validity of our training study and fills the gap in the ToM literature between the preschool children and middle childhood.

Our second prediction that, there would also be a significant improvement in the feedback without explanation condition from pre-test to post-test sessions compared to control condition, was also confirmed. In this condition, the feedback "Correct/Wrong" together with the correct answer was provided without providing further explanation. This result is important given the previous first-order ToM training studies that showed that 3- to 4-year-olds did not improve from pre-test to post-test session if they received "Correct/Wrong" feedback without explanation (Clements et al., 2000). An explanation for the improvement that we found derives from a previous computational cognitive modeling study. Arslan et al. (2013; 2017a) predicted that once children are able to attribute first-order false beliefs to an agent, they initially use a first-order reasoning strategy in second-order false belief tasks. The model predicts that later on, with repeated exposure to second-order ToM reasoning and with the help of feedback "Correct/Wrong", children can revise their first-order ToM strategy to one level higher, namely to a second-order ToM strategy. Note that Clements et al. (2000) did not provide the correct answer after the feedback "Correct/Wrong". Further research is needed with a training study of first-order false belief reasoning to test whether 3-year-old children's performance can be improved from pre-test to post-test when the correct answer is provided together with the feedback "Correct/Wrong" without any further explanations.

Our training study also shows other interesting and somewhat unexpected results. Children's performance from pre-test to post-test sessions also improved in the no feedback condition, and there was a marginally significant difference between the no feedback and control conditions ($B = 1.07$, $SE = 0.64$, $p = .09$). Considering Melot and Angeard's (2003) and Clements et al.'s (2000) findings that training 4-year-old children with first-order false belief tasks without providing any feedback and providing feedback without explanation did not improve children's performance, our results highlight the difference between first-order ToM and second-order ToM development. For the unexpected improvement in the no feedback condition, we surmise that exposing children to second-order ToM and asking them second-order false belief questions, together with the justification questions "Why?" helps children to reflect about their own judgments. Thus, asking justification questions helps children to revise their wrong strategy to a correct second-order ToM strategy. This argument needs to be tested with another training study in which children are trained on second-order ToM with no feedback, however, this time without asking the justification questions.

As can be seen from Figure 3.5 and Table 3.4, for all the conditions, children's scores on second-order false belief stories continued to improve 4 months after the pre-test session. Importantly, children who were in one of the three experimental conditions performed better at the follow-up session than the children who were in the control condition (73% correct in the feedback with explanation condition; 56% in the feedback without explanation condition; 68% in the no feedback condition; compared to 46% in the control condition). However, the greatest improvement at the follow-up session occurred in the feedback with explanation group, which significantly differed from the control condition. This result emphasizes the importance of further explanations in children's development of second-order ToM. The small improvement from pre-test to follow-up sessions in the control condition can be interpreted as the effect of children's natural development over 4 months combined with the effect of exposure to the second-order false belief tasks 9 times (3 stories each at pre-test, post-test, and follow-up sessions).

Moreover, we did not find any effect of children's verbal abilities and working memory score at pre-test on children's second-order false belief score. The insignificant effect of working memory is in line with Hasselhorn et al.'s (2005) study. We interpret the insignificant effect of working memory on children's second-order false belief score as a result of the simplicity of the counting span task. In order to succeed in the counting span task, children need just to count the blue shapes and make a list of numbers in their memory to report later. In contrast, we believe that a working memory task that needs more complex working memory strategies might predict children's second-order false belief scores. For example, a listening span task in which children are expected to first judge the truthfulness of each sentence by saying "Yes" or "No" and then have to recall the last word of all the sentences of a set told to them so far, in reverse order, can be a predictor of children's second-order ToM development. Further research is needed to verify this prediction.

3.4.2. Generalizability of the training effects

Before commenting on the generalizability of the training effects, it is worth discussing the differences between the three story types that we used in our study, namely 'Decoy gift', 'Three goals', and 'Three locations'. As we discussed in Subsection 3.2.3.1, 'Decoy gift' stories are less complex than the 'Three goals' and 'Three locations' stories because 'Decoy gift' stories have two possible answers compared to three possible answers in the other story types. In line with this explanation, in general, children's scores were higher in 'Decoy gift' stories. On the other hand, 'Three goals' stories and 'Three locations' stories both have three

possible answers for the second-order false belief question; indeed, there was no significant difference in children's scores between 'Three goals' and 'Three locations' stories at the pre-test session.

Note that we did not use 'Three locations' stories at the training sessions in order to test the generalizability of the training effect. In the experimental conditions, children's improvements in the stories that children trained on (i.e., 'Three goals' and 'Decoy gift' stories) did not significantly differ from their improvements in 'Three locations' stories from pre-test to post-test sessions (see Figure 3.6 and Table 3.3). These results together show the generalizability of the training effect.

Moreover, as can be seen in Table 3.4, we found an interesting and unexpected result about children's improvements in 'Three locations' stories from pre-test to follow-up sessions. In all conditions, children's improvements from pre-test to follow-up sessions in 'Three locations' stories were greater than their improvements in the stories that we used in the training sessions. More improvement in 'Three locations' over 4 months might be related to the linguistic structural differences of the second-order false belief questions in 'Three locations' stories compared to the second-order false belief questions in 'Decoy gift' and 'Three goals' stories. In 'Three locations' stories, the structure of the second-order false belief questions was in the form of a second-order embedding (e.g., "Where does Marieke think that Kevin will look for the chocolate?"). On the other hand, in 'Decoy gift' and 'Three goals' stories, the second-order false belief questions did not involve second-order embedding and were broken down into two pieces in order to facilitate children's comprehension (e.g., "The grandma asks 'What does Robert think that you will do today?' followed by the second-order false belief question 'What does dad say to grandma?'"). However, once children are more competent with second-order embedding, a second-order false belief question in the form of a second-order embedding might facilitate reasoning by delivering proper chunks ready for serialization (Hollebrandse & Roeper, 2014; de Villiers, Hobbs, & Hollebrandse, 2014). Further research is needed to investigate children's development over time in answering those two types of questions. Alternatively, in order to see whether children's improvements in the "Decoy gift" and "Three goals" stories will be as great as their improvements in the "Three locations" stories at the follow-up session, our training program can be replicated by asking the second-order false belief questions in second-order embedding form for all types of stories at the follow-up session.

In addition to the above-mentioned strengths of the design and the novelty of our findings, our study also has a number of limitations that should be acknowledged. First, the number of children in each condition is not large. Second, the variability in the socioeconomic status of the children was limited. Third, although

we used another second-order false belief story type to assess the generalizability of our training program, further research is needed to investigate whether children can generalize what they have learned from the second-order false belief training to their everyday social competence, such as idiom and irony understanding.

3.5. Conclusions

Our results showed that there are considerable improvements in children's scores from pre-test to post-test in the three experimental conditions: from 31% to 68% in the feedback with explanation condition; from 25% to 49% in the feedback without explanation condition; and from 33% to 55% in the no feedback condition, compared to a small improvement in the control condition (from 29% to 35%). The considerable improvements in the experimental conditions are not due to children's age, working memory, verbal abilities, or using a simple strategy instead of attributing second-order false beliefs.

Our findings that 5-year-olds' second-order false belief reasoning can be accelerated with different kinds of feedback both corroborate and contrast with the existing first-order ToM training studies. The improvements in the feedback with explanation condition are in line with the first-order ToM training studies that tested 3- to 5-year-olds and the advanced ToM studies that tested 9- to 10-year-olds. On the other hand, children's second-order false belief improvement in the feedback without explanation and in the no feedback conditions are contrary to the first-order ToM literature that showed that 3-year-olds' first-order ToM development could not be accelerated by training with feedback without explanations or without any feedback. Importantly, children can generalize the training effect to a story type on which they did not train and the training effect is stable over 4 months.

Based on our results, we suggest that children can be helped over the threshold to second-order false belief reasoning by the exposure to many stories and by asking them to reflect on second-order false belief questions, without providing explanations about their wrong answers.

Chapter 4: Syntactic Recursion Facilitates and Working Memory Predicts Recursive Theory of Mind

*In which we investigate the role of simple
and complex working memory strategies and
language on children's development of second-
order false belief reasoning.*



This chapter was previously published as:

Arslan, B., Hohenberger, A., & Verbrugge, R. (2017b). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PLOS ONE* 12(1): e0169510. doi: 10.1371/journal.pone.0169510

Abstract

In this study, we focus on the possible roles of second-order syntactic recursion and working memory in terms of simple and complex span tasks in the development of second-order false belief reasoning. We tested 89 Turkish children in two age groups, one younger (4;6 – 6;5 years) and one older (6;7 – 8;10 years). Although second-order syntactic recursion is significantly correlated with the second-order false belief task, results of ordinal logistic regressions revealed that the main predictor of second-order false belief reasoning is complex working memory span. Unlike simple working memory and second-order syntactic recursion tasks, the complex working memory task required processing information serially with additional reasoning demands that require complex working memory strategies. Based on our results, we propose that children's second-order theory of mind develops when they have efficient reasoning rules to process embedded beliefs serially, thus overcoming a possible serial processing bottleneck.

Keywords: second-order false belief reasoning; theory of mind; recursion; working memory; serial processing bottleneck

4.1. Introduction

Theory of mind (ToM) is the ability to understand that people have mental states, such as desires, beliefs, knowledge and intentions, and to realize that mental states of others might be different from one's own (Premack & Woodruff, 1978). Zero-order ToM reasoning concerns our real-life environment. For instance, if David thinks: "There is a newspaper on the table", he is applying zero-order reasoning. However, in daily life we are not just thinking about world facts. For example, David might think: "Jessica knows that there is a newspaper on the table". In this situation David engages in first-order ToM by making a first-order knowledge attribution to Jessica. In addition to first-order ToM, there are higher orders of ToM, such as David thinking, "Jack believes that Jessica knows that there is a newspaper on the table". This time, David is applying second-order recursion in the thought domain by attributing a first-order mental state to Jack.

First-order theory of mind has been found to be required for a number of simple social skills and competences. For example, children only start to be able to choose between informative ploys and deceptive ploys (such as removing tracks or adding false tracks) while hiding a toy when they are around the age of 4; their appropriate choices then correspond to first-order mental state attributions such as "now she will not know where the toy really is, but she will think it is under the cup where the tracks go" (Sodian, Taylor, Harris, Perner, 1992).

At the next step of development, second-order theory of mind has been found to be required for more advanced aspects of children's everyday social competence, such as idiom understanding (Caillies & Le Sourn-Bissaoui, 2013), which corresponds to second-order attributions like a hearer's reasoning "Peter is not really skating on thin ice, so the speaker wants me to think of a different meaning". As another concrete example, to successfully maintain a strategic lie, a lying child has to reason about what the listener knows about what the liar knows, requiring second-order theory of mind (Hsu & Cheung, 2013). Similarly, second-order theory of mind is a prerequisite for more complex moral judgments such as "the father knows that his daughter thinks that he will go to the pool, so he should really go there" (Perner, 1998). Finally, second-order theory of mind has been shown to be required for irony understanding ("although Oliver says 'You sure are a great scorer', Oliver doesn't really want Robert to believe that he is a great scorer") (Filippova & Astington, 2008).

Dennett (1978) argued that to have a theory of mind, a person has to be able to correctly attribute a false belief to someone else. Since then, verbal false belief tasks have become one of the most commonly applied tasks for testing theory of mind (Wimmer & Perner, 1983). The goal of the first-order false belief task is to examine whether children can attribute a false belief to another person in a given story

where the child knows the reality while the other person has a false belief about it. Similarly, the second-order false belief task examines whether children can correctly attribute to a person a false belief that that person has about another person's belief. While first-order false belief understanding develops around the age of four (Wellman, Cross, & Watson, 2001), second-order false belief understanding develops between the ages of five and seven (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994). The goal of this study is to investigate four- to eight-year-old children's development of second-order false belief reasoning.

One striking and much debated finding is that there is a delay between first- and second-order false belief reasoning in middle childhood. Why do children need some more years to pass second-order false belief tasks once they are able to pass first-order false belief tasks? The answer to this question is not entirely clear yet. Following the first-order ToM literature, two possible explanations have been proposed (see e.g., Miller, 2009, p. 751). The first explanation is related to a *conceptual change*: Children need to realize that mental states such as beliefs can have other beliefs and not just events in the world as their content (e.g., "John thinks that David believes that..."). The second explanation is related to the *complexity* of second-order ToM stories, in terms of the number of beliefs and their recursive organization. According to this explanation, it is the higher complexity of second-order ToM reasoning that adds further demands on working memory, as does the linguistic complexity of the stories and the questions, in comparison to first-order ToM tasks. Although we surmise that there might be a conceptual change of understanding that beliefs can have other beliefs as their content, in the current study we focus on the complexity explanation and aim to tease apart its components, namely executive functions and language.

For this purpose, for the first time in the literature, we focused on the role of working memory, together with the role of recursion in the language and thought domain on the same complexity and the same level of recursion, namely second-order.

In the following subsections of this Introduction, 'Working Memory and Theory of Mind' and 'Syntactic Recursion and Theory of Mind', we present relevant previous studies and provide theoretical explanations for the relations between working memory, language and second-order ToM.

4.1.1. Working memory and theory of mind

A number of other studies have shown that the development of executive function, more specifically working memory, influences the development of first-order ToM (e.g., (Gordon & Olson, 1998; Hughes, 1998; Keenan, Olson, & Marini,

1998; but see also Carlson, Moses, & Breton, 2002 for evidence that simple working memory is not sufficient without inhibition). On the other hand, to the best of our knowledge and according to Miller's (2009; 2012) extensive review, there are only two studies that focused on the role of working memory and second-order false belief reasoning in typically developing children and those studies yielded contradictory results.

The first one is Perner, Kain and Barchfeld's (2002) study with typically developing children and children at risk of ADHD. As a part of their study, to test executive function, they used forward and backward digit span. They found a significant positive relationship between the simple working memory task, i.e., the forward digit span, and the second-order false belief task. However, more recently, Hasselhorn, Mahler, and Grube (2005) also tested children around the age of six with a second-order false belief task and with a simple working memory task (a digit span task) and a non-word repetition task together with verbal ability tasks. Their results showed that the significant correlation between the simple working memory span score and children's second-order false belief score was no longer reliable when vocabulary knowledge and age were controlled for ($r(56) = .13, ns$).

What could be the role of working memory in the development of second-order false belief reasoning? It has been shown that working memory acts as a bottleneck (Borst, Taatgen, & van Rijn, 2010), meaning that people can only hold one chunk of information in working memory at a time. Given this restriction, we invoke the *serial processing bottleneck* hypothesis (Verbrugge, 2009).

Evidence for a working memory bottleneck in ToM reasoning comes from dual-task paradigms. The general idea of the dual-task paradigm is to find two different tasks and present them simultaneously, in order to compare performance with the two single tasks in which the participants have already performed well. If the performance of the first task decreases when it is presented concurrently with the second task, it can be inferred that the two tasks both require the same cognitive resource (Kahneman, 1973; Watanabe & Funahashi, 2015) and that that common resource acts as a bottleneck (Borst, Taatgen, & van Rijn, 2010; Salvucci & Taatgen, 2008). McKinnon and Moscovitch (2007) showed that young adults performed significantly worse on second-order ToM reasoning than on first-order ToM reasoning in a dual-task condition with a demanding secondary executive function task (i.e., the 2-back task). Participants in the single-task condition who only did second-order ToM reasoning did not show any loss in performance. This result was replicated in another study with young adults using a more advanced and naturalistic ToM task, namely the "Reading the Mind in the Eyes" task (Bull, Phillips, & Conway, 2007).

The *serial processing bottleneck* hypothesis can be seen as a procedural explanation of the cognitive process of serializing the hierarchical content of thought

into proper chunks along with their relations such that they fit easily through the processing bottleneck. Working memory constrains this process in terms of capacity and efficiency. The order of the processed chunks also reflects the reasoning steps children have to go through in order to solve a second-order false belief task. The *serial processing bottleneck* hypothesis is tantamount to a computational account under which second-order false belief can be conceived as a social-cognitive reasoning task employing a proper procedure – serialization – and a critical amount of mental resources – working memory – in order to cope with its nested structure.

Because working memory acts as a bottleneck, we propose that children who cannot pass second-order false belief tasks might have a lack of efficiency in reasoning when they have to serially process embedded beliefs. More specifically, when children try to answer second-order false belief questions, e.g., “Where does Mary think that John will look for the chocolate?” but have no efficient reasoning rule such as “if Mary didn’t see that John saw her hiding the chocolate, then she thinks that John thinks that the chocolate is still where he put it before, which is in the drawer”, more reasoning steps are needed to attribute a second-order false belief to Mary. A possible sequence of reasoning steps from the child’s perspective might be as follows: i) “John knows that the chocolate is in the toy box”, ii) “Mary does not know this”, iii) “Mary knows that John put the chocolate into the drawer before”, iv) “Mary thinks that John will look for the chocolate in the drawer”. In line with de Villiers, Hobbs, and Hollebrandse (2014), each reasoning step is in the form of a single-embedded sentence at the surface. In addition to de Villiers et al.’s study, the *serial processing bottleneck* hypothesis allows us to propose a possible explanation for their failures on children’s development of processing these rules to answer the second-order false belief question in terms of working memory.

If children do not have an efficient reasoning rule, they need to go through each of the reasoning steps i-iv, which occupies working memory temporarily. In order to proceed in reasoning, due to the working memory bottleneck, at each step, the information in working memory needs to be sent to long-term memory to be retrieved later, if necessary. Retrieving information from long-term memory also takes time and increases the odds of forgetting and of retrieving wrong information (Anderson & Schooler, 2000). Therefore, having more inefficient rules instead of one efficient rule means that the process is more prone to errors and takes more time (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Taatgen & Anderson, 2002). This view is consistent with research showing that children perform better in language comprehension tasks and cognitive tasks when they are given more time (Ling, Wong, & Diamong, 2015; van Rij, van Rijn, & Hendriks, 2010; Hendriks, van Rijn, Valkenier, 2006; Diamond, Kirkham, & Amso, 2002).

Once children have enough experience in applying these reasoning steps sequentially, they are combined to one efficient rule, repeated here for convenience: “if Mary didn’t see that John saw her hiding the chocolate, then she thinks that John thinks that the chocolate is still where he put it before, which is in the drawer” (see Taatgen & Lee, 2003 for the details of a mechanism that combines rules).

In order to investigate the *serial processing bottleneck* hypothesis, we looked at the relationships between a complex working memory task, a simple working memory span task, and second-order false belief reasoning. We did not use a dual-task paradigm because children have to be good at second-order false belief reasoning already in order to use a dual-task paradigm and our main focus is on children who are still developing second-order false belief reasoning. Although both simple and complex working memory tasks are related to working memory in a broader sense, they require different strategies. Unlike our simple working memory task that only requires building a representation of a list of words to be remembered, our complex working memory task requires processing information serially as well as cognitive control (Daneman & Carpenter, 1980; Unsworth & Engle, 2006). Therefore, we reason that children who have complex working memory strategies that can overcome the working memory bottleneck will be more successful in applying these in second-order false belief reasoning as well.

As a simple working memory span task, we used a word span task (WST) and as a complex working memory span task, we used a listening span task (LST). Considering the language-based nature of the listening span task, we used the word span task instead of the digit span task in order to keep the modality the same between the simple and complex span tasks. However, our results still can be compared with Perner et al.’s and Hasselhorn et al.’s study, because it has been shown that word span and digit span are closely related, $r = .65$, $p < .001$ (Henry, 2001) and have been grouped together with other span tasks which test the same component of working memory (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005).

We follow Carlson’s (2005) terminology, and refer to our simple word span task as a measure of pure working memory. Note that simple word span tasks have been referred to as short-term memory tasks as well (Cowan, Towse, Hamilton, Sauls, Elliot, Lacey et al., 2003). The details of the tasks are explained in the Methods section.

4.1.2 Syntactic recursion and theory of mind

Similar to the studies showing that language development contributes to the development of first-order ToM (de Villiers, Hobbs, Hollebrandse, 2014; Astington & Baird, 2005; de Villiers, 2007; de Villiers & de Villiers, 2014; Gleitman &

Papafragou, 2005; Milligan, Astington, & Dack, 2007; Walker & Murachver, 2012; but see Onishi & Baillargeon, 2005 for evidence of false belief understanding in preverbal infants), a number of studies have shown that language is important in children's development of second-order ToM.

For example, Hollebrandse, van Hout, and Hendriks (2014) compared six- to nine-year-olds' performance on a verbal and a low-verbal version of a second-order false belief task in order to investigate whether language in general helps children to pass second-order false belief tasks. They found that children's scores were lower in the low-verbal version of the second-order false belief task compared to the verbal version and concluded that language might support explicit reasoning about higher-order beliefs by facilitating tracking different beliefs. Similar to the findings of Hollebrandse, van Hout, and Hendriks (2011), Kuijper (2016) found that low-verbal second-order false belief tasks are also harder than high-verbal ones for children between six and twelve years of age who have been diagnosed with autism spectrum disorder or attention-deficit/hyperactivity disorder.

Lockl and Schneider (2007) found that, at the age of five, children's general language abilities (i.e., a combined score of sentence comprehension, morphological rule abilities and sentence memory) were strongly correlated with their second-order false belief reasoning. However, they stated that their data were not well suited to separate out the effects of syntactic and semantic abilities (p. 163). These studies indicate that explicit mental state language may support the development of second-order false belief reasoning. Yet, it is still not clear which aspect of language it is that helps.

Therefore, in this study, we investigate the possible relationship between syntactic recursion in the language domain and recursion in the thought domain on the same level of recursion, namely second-order.

The syntactic component of language is found to be related to first-order ToM in terms of its hierarchical embedding structure (de Villiers, Hobbs, & Hollebrandse, 2014; de Villiers, 2006; Hollebrandse, Hobbs, de Villiers, & Roeper, 2008; de Villiers, 2005; Hollebrandse & Roeper, 2014). Usually, first-order complement clauses, as shown below in Example (1) (adapted from de Villiers & de Villiers, 2014), have been used in the literature to investigate this relationship. Complement clauses such as "*that p*" may be used to express propositional attitudes (or opinions) towards some state *p* in the world (de Villiers, Hobbs, & Hollebrandse, 2014). They may be preceded by mental state verbs as in "Mary knows *that p*" or in "Mary believes *that p*" or by communication verbs as in "Mary said *that p*". Complement clauses can be used recursively, as shown below in Example (2). Moreover, complement clauses allow people to represent states that contrast with reality or with other people's mental states in terms of truth-value. Thus, while in Example (2), "Mary said that there was a flea in her cereal" might

be false, the whole sentence "John said that Mary said that there was a flea in her cereal" might be true.

- (1) First-order complement clause: "Mary said *that there was a spider in her cereal*. But it was just a raisin".
- (2) Second-order complement clause: "John said *that Mary said that there was a flea in her cereal*. But in fact, she said *that there was a spider in her cereal*".

Recently, de Villiers et al. (2014) argued that experience with truth-value contrasts in contexts with full tensed complement clauses such as Example (1) opens the door for children to pass first-order ToM tasks and to recognize syntactic recursion. Subsequently, understanding sentence recursion allows children to pass recursive ToM tasks. More specifically, they suggest that recursive complements in contexts in which the truth-values vary, such as in Example (2), are necessary for recursive false belief reasoning. However, they conclude that their predictions need to be explored further.

Like complement clauses, *relative clauses* can be used recursively; that is, besides first-order relative clauses, there are second-order relative clauses as well as even higher-order relative clauses. At each level of recursion they may refer to a different subject or object. However, unlike complement clauses, relative clauses do not involve propositional attitude verbs such as "knowing *that*" or "believing *that*" or communication verbs such as "saying *that*" and they do not involve truth-value contrasts. Table 4.1 shows examples of the progression of orders of recursion for ToM attributions and relative clauses. Unlike de Villiers et al. (2014), who focused on complement clauses, we used relative clauses, which allows us to specifically focus on the structural parallelism between second-order recursion in the language domain and in the thought domain by excluding the role of truth-value contrasts.

In the first-order domain, Hale and Tager-Flusberg (2003) demonstrated that preschoolers who were trained on first-order complement clauses improved their first-order ToM skills significantly while those trained on first-order relative clauses did not. On the other hand, Smith, Apperly, and White (2003) found a positive correlation between first-order relative clauses and first-order false belief tasks in children between the ages of 3 and 4. They concluded that first-order false belief reasoning might not be related to the specific structure of the complement clauses but to the broader category of embedded structures. Another developmental study with Turkish first-order relative clauses supports this positive relationship of first-order relative clauses and first-order ToM (Özoran, 2009).

Although the role of first-order relative clauses in first-order false belief reasoning has not been resolved yet, it is worthwhile to investigate the relationship between second-order relative clauses and second-order false belief reasoning.

Table 4.1. Examples of the progression of orders of recursion for theory of mind attributions and relative clauses.

Levels of recursion	Theory of mind	Relative clauses
Zero-order	The sheep is pushing a monkey.	The sheep is pushing a monkey.
First-order	You think that [the sheep is pushing a monkey].	You show me [the sheep that is pushing a monkey].
Second-order	You think that [the monkey thinks that [the sheep is pushing a monkey]].	You show me [the monkey that is pushing [a sheep that is pushing a monkey]].

This is because relative clauses share only the *syntactic* feature of embedded (meta-) representation with second-order false belief reasoning, in contrast to complement clauses, which additionally share semantic features with false belief, namely the fact that the main clause and the embedded complement clause have independent truth-values (Hale & Tager-Flusberg, 2003; , Smith, Apperly, & White, 2003; de Villiers & de Villiers, 2000). Since we are mainly interested in the structural and representational parallelism between recursion in the language and the thought domain, we chose a second-order relative clause comprehension task (REL_2) as a linguistic predictor of a second-order false belief task, which importantly concerns the same level of recursion. To the best of our knowledge, this is the first time that second-order relative clauses have been studied in relation to second-order false belief reasoning.

4.1.3. Predictions

1.

Based on the *serial processing bottleneck*, we predict that the relationship between the complex working memory task and the second-order false belief task will be stronger than the relationship between the simple working memory task and the second-order false belief task.
2.

Because the second-order false belief task and the second-order relative clause task share the same level of recursion, we expect a significant correlation between the two tasks.
- Therefore, we expect both the complex working memory task and the second-order relative clause task to figure importantly in the development of second-order false belief reasoning.

4.2. Method

4.2.1. Participants

Initially, a sample of 103 children between the ages three and eight was recruited from local kindergartens and primary schools in predominantly middle- and upper middle-class areas of Ankara, Turkey. Our study has been approved by Middle East Technical University (METU) Research Centre for Applied Ethics. A written parent approval form was obtained for every child that participated in the study. All children were monolingual Turkish native speakers. Three children left the study before it was completed. We excluded the youngest 11 children (range: 3;8 – 4;5) because they had very low scores in all tasks, indicating that the tasks were too hard for them in general. Thus, the results of 89 children were analyzed (37 female, $M_{age} = 6;7$ years, $SE = 0.13$, range: 4;5 – 8;10). Gender did not show any effect; therefore, the analyses were collapsed over gender. Considering the previous literature which indicates that second-order false belief reasoning starts to manifest itself between the ages five and seven (Perner & Wimmner, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994) and that Turkish children's development of first-order and second-order false belief reasoning shows a similar pattern with children in Western countries (Özoran, 2009; Girli & Tekin, 2010; Etel & Yagmurlu, 2014), we divided participants into the following two age groups: children younger than 6;6 were assigned to a *younger* age group and children older than 6;6 to an *older* age group:

- i)

Younger (4 - 6 years) $n = 41$; range = 4;6– 6;5; $M_{age} = 5;6$; $SE = 0.10$; 17 female,
- ii)

Older (6 – 8 years) $n = 48$; range = 6;7 – 8;10; $M_{age} = 7;6$; $SE = 0.11$; 20 female.

4.2.2. Design

A cross-sectional study design was used with age as a quasi-independent between-subjects variable. The same person tested all of the children in a quiet empty classroom at their school. For each child, all of the tests were completed in one session, which varied from 25 to 35 minutes.

All children participated in the following four tests in the following order: Word span task, second-order false belief tasks, second-order relative clause task, and listening span task. We wanted to keep some temporal distance between the two working memory tasks to prevent any interference effects. For this reason, we used one working memory task at the beginning and the other at the end of the session. Moreover, because we used children's second-order false belief task

scores as dependent variables, we used this task as a second task in our study, in order to prevent a possible fatigue effect. As Carlson and Moses (2001) argued, using a fixed order of tasks is standard practice in individual differences research and especially for interpreting correlations between tasks it is important that the task order remains the same for all participants.

Note that we also tested children with a pragmatic understanding task that we constructed. This task was presented after the second-order false belief tasks. In this task, children had to choose correct morphological case markers on an object – definite or indefinite – depending on whether the protagonists in a story had encountered the object before or not. However, this task did not correlate with any other task that we used in the experiment. For this reason, we do not present this task and its results here.

4.2.3. Materials and procedure

In this section, we present the materials and the procedures in the following order: second-order false belief task, second-order relative clause task, word span task, and listening span task. All of the stimuli can be found in S2 Materials.

4.2.3.1. Second-order false belief task (FBT_2)

This task consists of two different second-order false belief stories, namely the “Birthday Puppy Story” and the “Chocolate Bar Story”. Both stories were adapted from English to Turkish from Flobbe, Verbrugge, Hendriks, and Krämer’s (2008) study with the authors’ permission. These stories were told to the subjects while presenting Flobbe and colleagues’ drawings. Second-order embedding structures such as “Ayla thinks that Murat thinks that the chocolate is in the drawer” were not explicitly used in the stories. The order of stories was balanced.

While hearing a story, children were first asked a reality control question (“Where is the chocolate now?”), and a first-order ignorance question (“Does Murat know that Ayla has hidden the chocolate in the toy chest?”), as well as a linguistic control question (“Does Ayla know that Murat saw her hide the chocolate?”). The experimenter repeated the essential parts of the story and the control questions if a child gave a wrong answer for the control questions to make sure that children did not have any problems with remembering the stories and with the syntactic structure of double-embedded clauses. The upper limit for repeating the story and the control questions was three times (see Filipova & Astington, 2008; Benson,

Sabbagh, Carlson, & Zelazo, 2013 for repeating the control questions). All of the children gave correct answers to the control questions within that limit.

After the control questions, the children were asked (only once) a second-order false belief question: “Where does Ayla think that Murat will look for the chocolate?” and then (only once) a justification question: “Why does Ayla think that?”. In order to investigate the effects of our syntactic recursion and working memory tasks at different stages of second-order ToM development, we analyzed children’s judgments of the second-order false belief questions and justifications for their judgments separately (see Clements, Rustin, & McCallum, 2000 for an example of reporting judgment and justification answers separately). The rationale for analyzing both judgment and justification answers derives from a computational study showing that the operational demands of providing a justification are higher than the demands of making a judgment (Wahl & Spada, 2000). Therefore, it might be possible that the syntactic recursion and working memory have different effects on children’s judgments and justifications in second-order false belief tasks. A judgment score of 1 was given for a correct answer to a second-order false belief question, and a score of 0 was given for a wrong answer and for the answer “I don’t know”. Because we used two different stories, judgment scores could range from 0–2.

Children’s justifications for the second-order false belief task were coded based on the methods described by Perner and Wimmer (1985) and Sullivan et al. (1994). The categories were divided into two groups: correct and incorrect. If a child’s justification answer included the correct information that one character does or does not know about the other character’s history of exposure to relevant information, it was coded as correct. Otherwise, the justification was coded as incorrect (0 points). Correct justifications were divided into the following five mutually exclusive groups:

- a) *Explicit second-order reasoning*: The child embeds one character’s epistemic state in the other character’s mental state, for example, “Because she believes that Murat doesn’t know that the chocolate is in the box” (“Çünkü Murat’ın çikolatanın kutunun içinde olduğunu bilmediğini zannediyor”).
- b) *Implicit second-order reasoning*: Relevant information is embedded in one character’s epistemic state, for example, “Because she doesn’t know that Murat saw it” (“Çünkü Murat’ın gördüğünü bilmiyor”). Similar to Sullivan et al.’s (1994) study, we consider this statement to be second-order because of the role it plays in justifying a correct response.
- c) *Perceptive information*: Relevant information is embedded in one character’s perception, for example, “Because she didn’t see that Murat was looking through the window” (“Çünkü Murat’ın pencereden baktığını görmedi”).

- d) *Communicative information*: Information is mentioned that was communicated to the secondary character, for example, “Because she said she bought a ball” (“Çünkü top aldığını söylediği için”).
- e) *Location information*: The original location of the object is mentioned, for example, “Because Murat put it into the drawer before” (“Çünkü Murat çekmeceye koymuştu”).

As can be seen from the above-mentioned groups of justification answers, the sophistication of the answers differs. While the answers in the explicit and implicit second-order reasoning groups (a and b) include mental state words, the other three groups (c, d and e) do not include any mental state word. For these reasons, we gave 2 points for the answers in the explicit and implicit second-order reasoning categories and we gave 1 point for the answers in the other categories (see Filipova & Astington, 2008 for an example of a similar scoring procedure as ours, distinguishing different types of justifications based on the complexity of the answers). Because children were tested with two stories, the score range for their justifications was 0–4 in total.

Note that originally, we constructed three different versions of each second-order false belief story, in order to investigate the effect of three morphological evidential markers on the understanding of children’s second-order false belief reasoning: Neutral (present tense), –DI (past tense indicating direct perceptual evidence), –mİş (past tense indicating hearsay or inference). Only one of these versions was presented to each subject (see S2 Materials for details). Evidential markers encode the source of information and may therefore allow speakers and learners of evidential languages such as Turkish to take a positional perspective on a given propositional content – similar to propositional attitudes in false belief tasks (Aksu-Koç, 1988). However, because we did not find any significant difference between the three evidential conditions, we collapsed the data over them.

4.2.3.2. Second-order relative clause task (REL_2)

This task concerns the comprehension of relative clauses in Turkish and was adapted from Özge, Marinis, and Zeyrek’s (2010) first-order relative clause task with the authors’ permission. The questions and the drawings were modified to second-order relative clauses. Figure 4.1 demonstrates the drawings for one of the questions in this task. First, introductory pictures were shown to the participants in order to familiarize them with the animals in the actions by telling the name of the animals and the actions (e.g., “This is a pushing sheep, this is a looking monkey and this is a pushing monkey”). After that, the pictures representing the questions were shown one by one. The first and second rows of the picture were pointed

out in order to make clear that there were two separate lines of pictures by saying: “This is the first picture and this is the second picture”. In the practice session, the experimenter explained that the participants were required to point out the row with the animals corresponding to their answer. If they could not answer correctly in the practice session, the experimenter pointed out the correct animals and described their actions. However, no feedback was provided during the experimental session. The sentences were repeated up to 4 times. The critical positions for finding the correct answers were equally distributed across the drawings (3 times in the first row and 3 times in the second row) and between right (2 times), left (2 times), and central position (2 times). One practice item and 6 experimental items were used. A child’s total score for experimental items was minimally 0 and maximally 6.

There are other types of relative clauses, e.g., “You show me the sheep that a monkey is pushing that a sheep is pushing”. Because our aim is to examine the relationship between syntactic recursion and second-order false belief reasoning and not children’s different abilities in different types of relative clauses, we used relative clauses of the form “In which picture is there a sheep that is pushing a monkey that is pushing a sheep?”, which are straightforward to understand (Özge, Marinis, & Zeyrek, 2010). Children are exposed to recursion in relative clauses from an early age, for example in well-known nursery rhymes such as “This is the house that Jack built”, also cited in de Villiers and de Villiers (2014): “This is the maiden all forlorn, that milked the cow with the crumpled horn, that tossed the dog that worried the cat, that chased the rat that ate the cheese, that lay in the house that Jack built.”

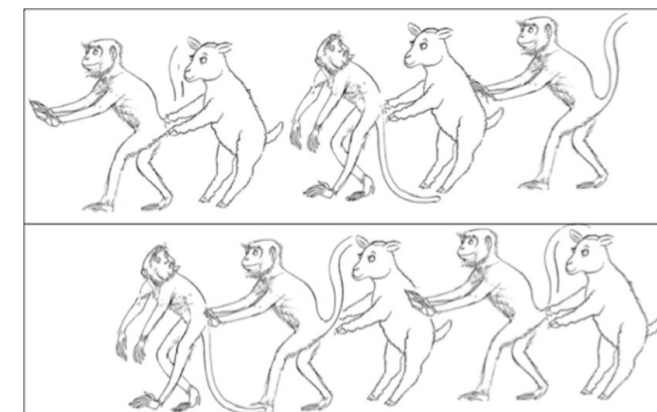


Figure 4.1. Picture used in the second-order relative clause task (REL_2). “In which picture is there a sheep that is pushing a monkey that is pushing a sheep?” Adapted from Özge et al. (2010) under a CC BY license, with permission from the authors.

Note that one might argue that children tend to interpret indirect recursion as conjunction reading. Thus, they might interpret “In which picture is there a sheep that is pushing a monkey that is pushing a sheep?” as “In which picture is there is a sheep that is pushing a monkey *and* a sheep?”. However, as you can see in Figure 4.1 and in S2 Materials, none of the pictures allow for such a conjunctive reading. Because all of our subjects pointed out three adjacent animals in one of the pictures throughout the task, we can say that this argument is ruled out.

4.2.3.3. Word span task (WST)

Children’s simple working memory span was tested using a Turkish version of the word span task (WST; Ünal, 2008). Monosyllabic Turkish words such as “saç”, “tuz” and “yurt” (hair, salt and country) were selected, considering their frequency in daily usage and ease of pronunciation. There were seven sets that corresponded to ascending levels of difficulty. Each level k contained three subsets of $k+1$ words each. At the first level, there were three subsets of two words each, and at the seventh (last) level, there were three subsets of eight words each. An example of the first level is: i) köşk – muz (manor – banana); ii) pil – üst (battery-upper); iii) buz – dört (ice – four), and an example of the seventh (last) level is: i) tam – bak – uç – göz – hal – boş – ek – yurt; ii) üç – kas – al – mülk – bir – tut – dil – kum; iii) bul – pek – on – fal – var – el – ses – genç. The words from these levels were read to the participants, starting from the first subset at the first level. After reading one subset (e.g., köşk – muz), the participant repeated the words in that order. If the participant could not correctly reproduce two out of three subsets at level $k+1$, the task was terminated and the level k was the score of the participant. Thus, in the analysis, a child’s word span range may vary between 0 and 7.

4.2.3.4. Listening span task (LST)

Children’s complex working memory span was tested using a Turkish version of the listening span task (LST; Ünal, 2008). The task consisted of sets of sentences read out to the participants one by one. There was a total of five collections, each of which consisted of six sets of sentences. The first collection contained six sets of two sentences each, the second collection contained six sets of three sentences each, and so forth, until the fifth collection, which contained six sets of six sentences each. An example of a 2-sentence set of LST is as follows: i) Muzlar bisiklete biner (“Bananas ride bicycles”); ii) Elimiz beş parmaklıdır (“Our hands

have five fingers”). The participants were expected to first judge the truthfulness of each sentence by saying “Yes” or “No”. Secondly, they had to recall the last word of all the sentences of a set told to them so far, in reverse order. After they gave an answer to the first sentence, the next sentence was told to them. For example, for the 2-sentence set, if the first sentence was “Muzlar bisiklete biner” (“Bananas ride bicycles”), the participants were required to say “Hayır; biner” (“No; bicycles”). After that, if the second sentence was “Elimiz beş parmaklıdır” (“Our hands have five fingers”), they were required to say “Evet; parmaklıdır, biner” (“Yes; fingers, bicycles”). If the participant made at most one mistake in a sentence collection, the subsequent sentence collection, which comprised one more sentence per set, was told to the participant. The score of the participants equaled the number of sentence collections in which they did not make more than one mistake. Thus, participants’ scores could range from 0-6.

4.3. Results

Our main goal was to investigate the role of language and working memory in the development of second-order ToM. In more detail, for the role of working memory, we aimed to test the *serial processing bottleneck* hypothesis by using both a complex and a simple working memory span task. Based on the *serial processing bottleneck* hypothesis, we predicted that the relationship between the complex working memory task and the second-order false belief task will be more salient than the relationship between the simple working memory task and the second-order false belief task. Because second-order relative clauses have the same level of recursion, we hypothesized that a child’s score on the second-order relative clause task could be a predictor of his or her second-order false belief scores.

The second-order false belief task (FBT_2) judgment scores ($W = 0.77, p < .001$) and the justification scores ($W = 0.782, p < .001$) were non-normally distributed. Because the data violated the normality assumption of ANOVA, a cumulative odds ordinal logistic regression with proportional odds was run (Agresti, 2013). The proportional odds and multicollinearity assumptions were satisfied. We first report the results for the development of tasks individually. We then report the bivariate and partial correlations among the tasks. Finally, we predict the FBT_2 judgment and FBT_2 justification scores by using ordinal logistic regression. Note that all the effect sizes (B) in the ordinal logistic regression are in terms of log odds.

4.3.1. Testing the differences between younger (4 – 6) and older (6 – 8) age groups

Table 4.2 presents the means together with the standard deviations of all the variables for each age group. Note that children between the ages 4;6 and 6;6 were assigned to the *younger* age group and between the ages 6;7 and 8;10 were assigned to the *older* age group:

The analyses showed that while children in the older (6 – 8 years) group outperformed children in the younger (4 – 6 years) group in the FBT_2 judgment scores ($B = 1.29$, $SE = 0.42$, $p = .002$), there was no significant difference in the FBT_2 justification scores between the younger and older age groups ($B = 0.35$, $SE = 0.39$, $p = .38$). Moreover, children in the older group outperformed children in the younger group for the REL_2 task, ($B = 0.88$, $SE = 0.39$, $p = .03$), for the WST task, ($B = 1.69$, $SE = 0.45$, $p < .001$), and for the LST task, ($B = 1.87$, $SE = 0.47$, $p < .001$).

Table 4.3 shows the number of participants and percentages (in parentheses) for each second-order false belief (FBT_2) judgment score (0–2) and justification score (0–4). Consistent with the literature (Miller, 2012), judgment scores for the second-order false belief question were a bit higher for the ‘Birthday

Table 4.2. Descriptive statistics of the four tasks administered to each age group

Tasks	Younger (4 – 6 years)	Older (6 – 8 years)
	Mean (SD)	Mean (SD)
False belief judgment (Range 0 – 2)	1.04 (0.82)	1.54 (0.58)
False belief justification (Range 0 – 4)	1.10 (1.30)	1.29 (1.27)
Relative clause task (REL_2) (Range 0 – 6)	1.15 (1.49)	1.92 (1.78)
Word span task (WST) (Range 0 – 8)	4.05 (0.80)	4.81 (0.87)
Listening span task (LST) (Range 0 – 6)	0.34 (0.69)	1.13 (0.98)

Table 4.3. Number of participants and percentage (in parentheses) of each second-order false belief (FBT_2) score

Age Group	FBT_2 judgment score			FBT_2 justification score				
	0	1	2	0	1	2	3	4
Younger (4 – 6)	13 (32%)	14 (34%)	14 (34%)	20 (49%)	5 (12%)	12 (29%)	0 (0%)	4 (10%)
Older (6 – 8)	2 (4%)	18 (38%)	28 (58%)	18 (37%)	10 (21%)	11 (23%)	6 (13%)	3 (6%)

Puppy’ story than for the ‘Chocolate Bar’ story for both the younger age group ($M_{score.chocolate\ bar} = 0.46$, $SD = 0.50$; $M_{score.birthday\ puppy} = 0.56$, $SD = 0.50$) and the older age group ($M_{score.chocolate\ bar} = 0.69$, $SD = 0.47$; $M_{score.birthday\ puppy} = 0.85$, $SD = 0.36$).

The detailed results about the frequency and percentage of each type of justification answer are shown in Table 4.4. As can be seen from Table 4.4, children’s correct justification answers mostly involved implicit second-order answers (e.g., “Because she doesn’t know that Murat saw it”) for both age groups. Moreover, while there were two children in the older age group who gave explicit second-order justification answers (e.g., “Because she believes that Murat doesn’t know that the chocolate is in the box”), none of the children in the younger age group gave any explicit second-order answers. However, children’s justification answers in the older age group (6 – 8) are clearly not at the ceiling and probably continue to develop after the age of 8.

4.3.2. Bivariate and partial correlations

To inspect the interrelationships among the four tasks, we conducted bivariate and partial correlations. Table 4.5 shows the bivariate correlation coefficients

Table 4.4. Frequency (Freq.) and percentage (%) of each type of justification answers

Story type	Justification type	Younger (4 – 6 years)		Older (6 – 8 years)	
		Frequency	%	Frequency	%
Chocolate Bar story	Explicit second-order	0	0	1	2
	Implicit second-order	13	32	10	21
	Perceptive information	1	2	9	19
	Communicative information	0	0	0	0
	Location information	2	5	6	12
	Wrong answers	25	61	22	46
Total		41	100	48	100
Birthday Puppy story	Explicit second-order	0	0	1	2
	Implicit second-order	7	17	10	21
	Perceptive information	0	0	0	0
	Communicative information	2	5	3	6
	Location information	0	0	0	0
	Wrong answers	32	78	34	71
Total		41	100	48	100

(Spearman's r_s) and Table 6 shows the partial correlation coefficients (Spearman's r_s) for the younger (4 – 6 years) and older (6 – 8 years) groups. We used age (in months) as a control variable.

As can be seen from Table 4.5, for the younger group (4 – 6), there is a significant correlation between the FBT_2 judgment score and REL_2 ($r_s = 0.33$, $p = .03$), and a significant correlation between the FBT_2 judgment score and LST ($r_s = .51$, $p < .001$). However, as shown in Table 4.6, correlations between the FBT_2 judgment score and REL_2 become insignificant when we control for LST ($r_s = -.001$, $p = .99$) but remain significant when we control for age ($r_s = .31$, $p = .04$) and for WST ($r_s = .30$, $p = .04$). On the contrary, the correlation between the FBT_2 judgment score and LST remains significant when we control for age ($r_s = .50$, $p < .001$), REL_2 ($r_s = .41$, $p = .006$) and WST ($r_s = .52$, $p < .001$). For the older group (6 – 8), none of the tasks show significant correlations with the FBT_2 judgment score. The lack of significant relationships between LST and older children's FBT_2 judgment scores (range 0–2) is due to the fact that the older children already performed well in providing judgment answers, so there is a lack of enough variation in the data.

Table 4.5. Bivariate correlation coefficients (Spearman's r_s) for the younger (4 – 6 years) and older (6 – 8 years) age groups

		Younger group (4 – 6 years)					Older group (6 – 8 years)				
		1 ^a	2	3	4	5	1	2	3	4	5
FBT_2 Judgment (range: 0 – 2)	1. Age (in months)	-	-	-	-	-	-	-	-	-	-
	2. Judgment	.27	-	-	-	-	.13	-	-	-	-
	3. Relative clause task (REL_2)	.14	.33*	-	-	-	.001	.09	-	-	-
	4. Word span task (WST)	.15	.15	.32*	-	-	.15	.13	.42**	-	-
	5. Listening span task (LST)	.10	.51***	.66***	-.05	-	.15	.17	.41**	.38**	-
FBT_2 Justification (range: 0 – 4)	1. Age (in months)	-	-	-	-	-	-	-	-	-	-
	2. Justification	.14	-	-	-	-	.21	-	-	-	-
	3. Relative clause task (REL_2)	.14	.48**	--	-	-	.001	.22	-	-	-
	4. Word span task (WST)	.15	.29	.32*	-	-	.15	.15	.42**	-	-
	5. Listening span task (LST)	.10	.59***	.66***	-.05	-	.21	.37**	.41**	.38**	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

^a The numbers in this row are used as abbreviations for the age and the tasks that were enumerated in the second column of this table.

Similar to the FBT_2 judgments scores, for the younger age group, there is a significant correlation between the FBT_2 justification score and REL_2 ($r_s = .48$, $p = .001$), and a significant correlation between the FBT_2 justification score and LST ($r_s = .59$, $p < .001$). Moreover, there is a marginally significant correlation between the FBT_2 justification score and WST ($r_s = .29$, $p = .07$). As partial correlations reveal in Table 4.6, only the correlation between the FBT_2 justification score and LST remains significant when we control for age ($r_s = .58$, $p < .001$), REL_2 ($r_s = .41$, $p = .006$) and it even increases somewhat when we control for WST ($r_s = .63$, $p < .001$), previewing the results of the subsequent regression analyses.

For the FBT_2 justification scores in the older age group, there is only a significant correlation between the FBT_2 justification score and LST ($r_s = 0.37$, $p = .009$). As shown in Table 4.6, the correlation between the FBT_2 judgment score and LST remains significant when we control for age ($r_s = .34$, $p = .01$), REL_2 ($r_s = .32$, $p = .02$) and WST ($r_s = .35$, $p < .01$).

Moreover, for both FBT_2 judgment and justification scores, we found a significant high correlation between LST and REL_2 ($r_s = .66$, $p < .001$). These correlations are still significant when we control for age and the other tasks that we used in this study, indicating that LST and REL_2 share a considerable amount of variance.

Table 4.6. Partial correlation coefficients (Spearman's r_s) for the younger (4 – 6 years) and older (6 – 8 years) groups

		Younger group (4 – 6 years)					Older group (6 – 8 years)			
		1 ^a	2	3	4		1	2	3	4
FBT_2 Judgment	1. Age (in months)	-	-	-	-	-	-	-	-	-
	2. Relative clause task (REL_2)	.31* ^b	-	.30*	-.001	.09	-	-	.03	.02
	3. Word span task (WST)	.12	.05	-	.20	.12	.11	-	-	.08
	4. Listening span task (LST)	.50***	.41**	.52***	-	.15	.14	.13	-	-
FBT_2 Justification	1. Age (in months)	-	-	-	-	-	-	-	-	-
	2. Relative clause task (REL_2)	.47***	-	.43**	.16	.22	-	-	.08	-
	3. Word span task (WST)	.27	.16	-	.39**	.12	.06	-	-	.004
	4. Listening span task (LST)	.58***	.41**	.63***	-	.34*	.32*	.35*	-	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

^a The numbers in this row are used as abbreviations for the age and the tasks that were enumerated in the second column of this table.

^b The partial correlations show the correlation between a variable in a row and judgment/justification score when a variable in a column is controlled for. For example, (.31*) shows the partial correlation between REL_2 and judgment score when age (in months) is controlled for.

4.3.3. Predicting the second-order false belief task score (FBT₂)

The results of the regression models that best predict second-order false belief judgment and justification scores for both the younger (4 – 6 years) and the older (6 – 8 years) age groups are presented in Table 7. We proceeded as follows with our model construction for the younger group: In model 1, we entered the control variable “age in months” in order to account for any more fine-grained age differences within the younger age group, as well as the second-order relative clause task (REL₂), and the complex working memory task (LST), because we had found significant bivariate correlations between the FBT₂ judgment score and these two tasks. Moreover, even though we did not find strong significant correlations between the simple working memory task (WST) and FBT₂ judgment and justification scores, we kept WST in our models because we had a specific prediction that LST is more related to second-order false belief reasoning than WST is, based on the *serial processing bottleneck* hypothesis. In Model 1, the effects of age in months ($B = 0.05$, $SE = 0.04$, $p = .22$), REL₂ ($B = -0.51$, $SE = 0.36$, $p = .16$), and WST ($B = 0.68$, $SE = 0.46$, $p = .14$) were insignificant; only the effect of LST was significant ($B = 3.44$, $SE = 1.35$, $p = .01$). As shown in Table 5 and Table 6, considering the strong correlations between REL₂ and LST, and based on the result of Model 1 that REL₂ is not significant, we constructed Model 2 by excluding REL₂ from Model 1. As can be seen from Table 7, the effect of LST on the FBT₂ judgment score is significant ($B = 2.22$, $SE = 0.82$, $p = .007$), and the effect of WST is insignificant ($B = 0.39$, $SE = 0.40$, $p = .33$). The model comparison of Model 1 (Akaike Information Criterion (AIC) = 82.60) and Model 2 (AIC = 82.77) was less than 2. Because a simpler model is preferred over a more complex one, we presented the results of Model 2 in Table 7.

We followed the same procedure as explained above for the prediction of the younger age group's FBT₂ justification scores. We entered the control variable “age in months”, REL₂, LST, and WST into Model 3. In Model 3, the effects of age ($B = -0.006$, $SE = 0.04$, $p = .89$), and REL₂ ($B = -0.10$, $SE = 0.30$, $p = .75$) were insignificant, and the effects of LST ($B = 2.05$, $SE = 0.73$, $p = .005$), and WST ($B = 1.18$, $SE = 0.53$, $p = .03$) were significant. Based on the results of Model 3, we constructed Model 4 by excluding REL₂ from Model 3. As can be seen from Table 7, the effects of LST ($B = 1.91$, $SE = 0.57$, $p < .001$) and WST ($B = 1.12$, $SE = 0.49$, $p = .02$) on the FBT₂ justification score are significant. Because Model 4 (AIC = 90.04) is a simpler model than Model 3 (AIC = 91.94), we presented the results of Model 4 in Table 7.

For the older group (6 – 8 years), we had found a significant correlation only between LST and the FBT₂ justification score, and for them, none of the tasks were significantly correlated with the FBT₂ judgment score due to the lack of

variation. Although none of the tasks were significantly correlated with the FBT₂ judgment score, because the younger group's (4 – 6 years) final model to predict the judgment score includes LST and WST, we constructed Model 5 to predict the older group's FBT₂ judgment score by adding LST and WST. As shown in Table 4.7, in line with the lack of significant correlations, both LST and WST's effects were insignificant in the model.

Similar to the previous procedures, in order to predict older children's FBT₂ justification scores, we constructed Model 6 by entering the control variable “age in months”, WST, and LST. The effect of LST is significant when we control for age ($B = 0.79$, $SE = 0.33$, $p = .02$), and the effect of WST is insignificant ($B = -0.05$, $SE = 0.34$, $p = .88$).

These results suggest that the main predictor of second-order false belief reasoning is not syntactic recursion and word span task but complex working memory.

Table 4.7. Predicting second-order false belief task judgment and justifications scores for both younger (4 – 6) and older (6 – 8) age groups

	Variable	Younger group (4 – 6 years)				Older group (6 – 8 years)			
		B	SE	t	p	B	SE	t	p
FBT ₂ Judgment	Age (in months)	0.06	0.04	1.48	.138	0.03	0.03	0.82	.41
	Word span task (WST)	0.39	0.40	0.97	0.33	0.08	0.36	0.21	.83
	Listening span task (LST)	2.16	0.82	2.67	.008	0.27	0.32	0.83	.41
FBT ₂ Justification	Age (in months)	-0.005	0.04	-0.10	.92	0.03	0.03	1.02	.31
	Word span task (WST)	1.12	0.49	2.29	.02	-0.05	0.34	-0.16	.88
	Listening span task (LST)	1.91	0.57	3.35	<.001	0.79	0.33	2.42	.02

4.4. Discussion

The main goal of this study was to investigate the role of syntactic recursion and working memory in the development of second-order false belief reasoning as well as to provide a procedural account for the role of working memory. In order to focus on the different stages of children's development of second-order false belief reasoning, we have run separate analyses of children's judgments for the second-order false belief question and justifications for their judgments. Our results showed that the main and strongest predictor of the development of the second-order false belief reasoning is the complex working memory span (LST).

Let us first discuss the results related to the simple and the complex working memory tasks. Considering the literature that we discussed in the Introduction, our finding that there is no significant correlation between our simple working memory task and the second-order false belief judgement scores for both younger (4 – 6 years) and older (6 – 8 years) age groups is in line with Hasselhorn et al.'s study (2005) that found no significant correlation between the simple working memory task and the second-order false belief task when the effects of verbal ability tasks and age were controlled for. Furthermore, our findings are also consistent with the previous literature that shows that complex working memory tasks are better predictors of measures of general intelligence than simple working memory tasks (Ackerman, Beier, & Boyle, 2005; Engle, Tuholski, Laughlin, & Conway, 1999). However, as shown in Table 4.7, in addition to the highly significant effect of the complex working memory task, the simple working memory task explains significant variation in younger children's (4 – 6) justification answers. This significant correlation of the simple working memory task disappears for older children and only the complex working memory task is able explain the variance in children's justification answers. As we predicted, we found a significant correlation between our complex working memory task and the second-order false belief judgment and justification scores, even when we controlled for the simple working memory task, the second-order relative clause task and age in months – in the younger age group. Moreover, for the older age group, for the complex working memory task, we found that it only significantly predicts the second-order false belief justification score, not the second-order judgment score. The reason is that the judgment scores do not differ much among the older children, while their justifications still do. Thus, justifications seem to be a more sensitive variable for older children in the sense that they provide a finer distinction in their second-order reasoning abilities. While older children can give correct second-order false belief answers, their development still continues in terms of their justification abilities.

Now, let us discuss our results in terms of one of our main goals, namely, testing the *serial processing bottleneck* hypothesis (Verbrugge, 2009). Why is the complex working memory task more important than the simple working memory task in predicting children's performance on second-order false belief reasoning, in terms of both judgment and justification scores? The *serial processing bottleneck* hypothesis predicts that the difficulty of passing a second-order false belief task is not just related to holding the different beliefs in mind but also to serially processing them. In order to test this prediction, we have used both a simple working memory task (WST) that requires just holding the information in mind and a complex working memory task (LST) that requires not only holding information in mind but also processing that information serially, as well

as additional reasoning demands that require complex working memory strategies. We argue that these differences between the complex working memory and simple working memory tasks could be the reason why the simple working memory task cannot explain enough variation of children's performance on second-order false belief reasoning. Two subtasks appear to be required for successful second-order false belief reasoning: (1) keeping in mind the two separate beliefs (e.g., of John and Mary) and (2) mapping their nested, recursive structure onto the appropriate sequential order: Mary's belief that John believes that p, such that they can pass the *serial processing bottleneck* smoothly. A few young children and many older children overcome the *serial processing bottleneck* by means of their complex working memory strategies, which are necessary for both the complex working memory task and the second-order false belief reasoning.

It is important to discuss some additional challenges of the Turkish version of the complex working memory task that we used, namely the listening span task. First, Turkish is a verb-final language – hence, the final word of the sentence, which is the critical word to be reported in the listening span task, may be a verb. Verbs and nouns have different semantic and computational loads and may therefore not be memorized equally well in the listening span task. Second, because in Turkish, the present form of the verb takes the suffixes –er, –ar, –ir, –ür, –ur (depending on vowel harmony) for positive sentences while it takes the suffixes –maz, –mez for negative sentences, an additional challenge of the task for children in our study was to repeat the last word of the sentence when the sentence was false and they had to say “Hayır” (“No”). So, participants had to inhibit the negative form of the final verb, e.g., for the sentence “Muzlar bisiklete biner” (“Bananas ride bicycles”) they should not say “binmez” (“they don't ride”) but “biner” (“they ride”). This additional load in inhibition, due to the way the Turkish morphological system works, may cause the listening span task results to be somewhat different for Turkish children than for English-speaking children. Moreover, similar to what is argued in Carlson, Moses, and Breton (2002) and Moses, Carlson, and Sabbagh (2005) for the relation between first-order ToM and executive function, the additional inhibition demands in the Turkish version of the listening span task might be one of the reasons for its predictive power of the development of second-order ToM reasoning in our Turkish sample. Note that there were only 9 children (out of 41) in the younger group who scored more than 0 in the listening span task. However, for those 9 children, the listening span task score still significantly predicts the second-order false belief score in the ordinal logistic regression models. Further cross-linguistic studies are needed to unravel possible developmental differences in the listening span task between children speaking typologically different languages.

Now we will focus on the results about the role of syntactic recursion in the development of second-order ToM that we tested by constructing a second-order relative clause task (REL_2). As mentioned before, there is no consensus on the relationship between first-order relative clauses and first-order ToM (Hale & Tagger-Flusberg, 2003; Smith, Apperly, & White, 2003; Özoran, 2009). In line with those studies that showed a positive relationship between the two, in the younger age group, we found a significant relationship between the second-order false belief judgment score and our syntactic recursion task, a relationship of $r_s = .31$ ($p = .04$) when age was controlled for. Moreover, we found a significant relationship between the second-order false belief justification score and our syntactic recursion task, $r_s = .47$ ($p < .001$) when age was controlled for. This two-fold positive relationship supports the view that a purely structural parallel between the linguistic realm and reasoning (de Villiers, 2007; de Villiers, 2005) does hold for the development of second-order ToM. However, this relationship completely disappeared when we controlled for the complex working memory task. This loss is due to the very strong correlation between the complex working memory task and the second-order relative clause task ($r_s = .66$, $p < .001$ for the younger group, and $r_s = .41$, $p = .004$ for the older group). This common variance of both tasks is again shared with the second-order false belief score. These strong mutual correlations are consistent with the hypothesis that the *serial processing bottleneck* seems to strongly affect all three tasks, at younger and older ages.

Our findings appear to be of interest to both language and memory researchers. The findings indicate, overall, that complex working memory strategies play a larger role in second-order ToM reasoning than syntactic recursion. However, given the strong overlap between the second-order relative clause task and the complex working memory task in both age groups, as shown in Table 4.5, it seems plausible that they both require similar complex working memory strategies that also facilitate second-order false belief reasoning. This similarity may give a hint at a possible convergence between the language and the memory explanations. As for the language explanation, hierarchical, syntactic embedding may be just the right representational tool to aid in the serialization process. The propositions are lexically selected by matrix verbs (e.g., “say” or “think”) as in embedded complement clauses (e.g., “John said that Mary said that there was a flea in her cereal. But in fact, she said that there was a spider in her cereal”), or dependent on a head noun (e.g., “the sheep”) as in embedded relative clauses (e.g., “Show me the sheep that is pushing a monkey that is pushing a sheep”). Furthermore, they are clearly demarcated and introduced by functional heads (“think that”; “the sheep that”). Thus, they are delivered in proper chunks ready for serializing them and passing smoothly through the *serial processing bottleneck* where central processes of interpretation take place. It is this chunking that may facilitate reasoning

about the various beliefs (John’s and Mary’s, as pointed out above) in correct order. Importantly, horizontal, serial order directly follows from vertical, hierarchical structure: What is higher in the structural representation precedes in the linear string. As Hollebrandse and Roeper (2014) state: “Recursion in grammar involves a translation between a hierarchical into a linear structure”.

The fact that the complex working memory task and not the second-order relative clause task is a better predictor for the second-order false belief reasoning may be due to the additional reasoning component in both our complex working memory task and the second-order false belief task, which is lacking in our second-order relative clause task. For example, to be able to pass the second-order relative clause task, one should parse the question “In which picture is there a sheep that is pushing a monkey that is pushing a sheep?” to obtain the meaning and select the proper picture that is the correct answer. Because the question includes second-order recursion, it also needs serial processing of information. However, it is possible to check the intermediate steps of the embedded parts of the sentence visually while parsing the sentence from the presented figure (Figure 4.1), which reduces the demands of working memory.

On the other hand, when answering the second-order false belief question “Where does Mary think that John will look for the chocolate?”, one is not only parsing the sentence to get the meaning, one also reasons about the question to come up with an answer. To be able to give a correct answer, one has to reason about the contradictory knowledge of Mary and John based on the reasoning rules, such as “Mary did not see John saw her hiding the chocolate, so she thinks that John thinks that the chocolate is still where he put it before, which is in the drawer, and therefore Mary thinks that John thinks that the chocolate in the drawer”. To achieve this reasoning, one should have efficient working memory strategies to overcome the *serial processing bottleneck*. Similarly, LST also requires an additional reasoning component beyond just holding in mind the to-be-remembered items and parsing the sentence, which is judging the truth value of the sentence. In addition, in our Turkish version of the LST, participants had to suppress the negative morphological marker of the verb when the truth value was negative, as discussed above.

As mentioned before, de Villiers et al. (2014) argued that the truth contrasts in contexts with first-order complement clauses (“Mary said that there was a spider in her cereal. But it was just a raisin”) open the door for children to pass first-order false belief tasks and to recognize syntactic recursion. They further argue that, subsequently, understanding sentence recursion in contexts with second-order complement clauses allows children to pass higher-order theory of mind tasks (e.g., second-order false belief tasks). In addition to de Villiers et al.’s (2014) argument that the truth contrasts might be an important stepping-stone in children’s

understanding of sentence recursion which facilitates recursive false belief reasoning, we propose a general explanation for the development of children's second-order false belief reasoning.

We surmise that children start to pass first-order false belief tasks when they learn to overcome the *serial processing bottleneck* by constructing more efficient reasoning rules to be able to attribute a false belief to another agent (first-order ToM) than applying the most salient reasoning strategy, that is, zero-order reasoning. Similarly, children pass second-order false belief tasks when they again learn to overcome the *serial processing bottleneck*, but this time by constructing efficient reasoning rules for second-order ToM reasoning. Our theory can be tested by adapting the standard second-order false belief tasks so that it is possible to derive children's level of reasoning (i.e., zero-order, first-order, second-order) from their answers to the second-order ToM questions (see Hollebrandse, Hobbs, de Villiers, & Roeper, 2008 for an example of a 'Bake Sale' story, in which a child can answer a second-order question with reference to three different objects, which correspond one-to-one to the three levels of ToM reasoning). We expect that children around the ages of 5 and 6 who cannot pass second-order ToM tasks will give mostly first-order answers instead of zero-order answers (reality bias). The *serial processing bottleneck* hypothesis also provides a procedural explanation of de Villiers et al.'s (2014) following argument about children's failure in second-order recursive structures (i.e., "Mary believes that John thinks that ..."): "In both complementation and false belief reasoning, children first treat 2-level embedding as 1-level of structure. It is as if one piece of the hierarchy is flattened, or skipped over in parsing." (p. 239).

We may generalize children's failures at first-order and second-order false belief reasoning by saying that children's incorrect answers are typically one order below the target order of false belief reasoning. Consistent with Miller's account in terms of complexity (2009, p. 751), our parallel construal of first- and second-order ToM reasoning as well as the similarity in the patterns of failure may indicate that there is a common process underlying the development of first- and second-order ToM reasoning. If cognitive control over competing representations is gained and the nested structure of these representations can be serialized appropriately, children are capable of second-order ToM reasoning. Although we have discussed our results in terms of the complexity account, our results do not exclude the possibility that children's recursive language abilities and complex working memory strategies may also contribute to a possible conceptual change that beliefs can be recursive.

Also note that we presented the tasks in the following fixed order: 1) simple working memory task; 2) second-order false belief task; 3) second-order relative clause task; 4) complex working memory task. Although there is no a priori

reason that this particular presentation of the tasks might have produced the particular effects (for a similar case, see Beck, Riggs, & Gorniak, 2009), future research is needed to rule out any effect of the order of the tasks.

4.5. Conclusions and future directions

As we predicted in the Subsection 4.1.3. "Predictions" of this chapter, there is a significant relation between the complex working memory task and the second-order false belief task and this relation is stronger than the relation between the simple working memory task and the second-order false belief task. Moreover, as we predicted, younger children's (4 – 6) double-embedded relative task score is significantly correlated with their second-order false belief task score. However, our study shows that the main predictor of the development of second-order theory of mind (ToM) is the complex working memory task for both children's judgment and justification answers for the second-order false belief question. Our study also shows that syntactic recursion and complex working memory measures are inter-related, suggesting common underlying capacities and processes. Based on these results, we propose that children's second-order ToM develops when they are able to apply efficient reasoning rules to process embedded beliefs serially, thus overcoming the *serial processing bottleneck*.

To further test the *serial processing bottleneck* hypothesis, future research is needed, possibly with a training study in which the children are trained with a complex working memory span task while a simple working memory task is used in a control group. In this way, the effect of complex working memory strategies on second-order ToM reasoning can be observed. Moreover, to test whether children's second-order false belief reasoning is supported more by second-order complement tasks, as argued by de Villiers et al. (2014), or by complex memory tasks, one could also design a training study in which children on the brink of second-order ToM are subjected to training regimes consisting of second-order 'memory of complement' tasks (condition 1) or various complex working memory tasks (condition 2) and compare their improvements on second-order false belief tasks. Furthermore, to test whether the relationship between syntactic recursion and second-order false belief reasoning holds exclusively for recursion on the clause level or for recursion of any constituent, possessive recursions (as in "Mary's friend's dress") might be used (Pérez-Leroux, Castilla-Earls, Bejar, & Massam, 2012). In addition to testing these hypotheses with behavioral data, constructing computational cognitive models by using cognitive architectures is a promising line of research (e.g., van Rij, van Rijn, & Hendriks, 2010).

Chapter 5: The Role of Simple and Complex Working Memory Strategies in the Development of First-order False Belief Reasoning: A Computational Model of Transfer of Skills

*In which we investigate the far transfer
from simple and complex working memory
strategies to first-order false belief reasoning by
constructing computational cognitive models
using the cognitive architecture PRIMs.*



This chapter was previously published as:

Arslan, B., Wierda, S., Taatgen, N., & Verbrugge, R. (2015b). The role of simple and complex working memory strategies in the development of first-order false belief reasoning: A computational model of transfer of skills. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.), *Proceedings of the 13th International Conference on Cognitive Modeling*, (pp.100-105). Groningen: University of Groningen.

Abstract

In their fourth year, most children start to understand that someone else might have a false belief, which is different from the reality that the children know. The most studied experimental task to test this development is called the first-order false belief task. What kind of prior cognitive skills help children to pass the false belief task? There are hundreds of correlational studies that have shown that language and executive functions (such as inhibition and working memory) play a role. Moreover, several training studies have shown the importance of language and inhibition in the development of false belief reasoning. However, to the best of our knowledge there has been no training study (with normally developing children) to investigate the role of working memory strategies in the development of false belief reasoning.

We present here a computational cognitive model to investigate transfer from working memory strategies to false belief reasoning. For this reason, in addition to the false belief task, we constructed two tasks that children encounter in their daily life: a pencil task (simple working memory) and a marble task (complex working memory). Our simulation results confirm our hypothesis that there is more transfer from the marble task to the first-order false belief task than from the pencil task to the first-order false belief task, because of the more complex working memory strategies that appear to be necessary in the false belief task. The results of our simulations suggest conceptual predictions to be tested experimentally.

Keywords: theory of mind; false belief reasoning; working memory; transfer; cognitive modeling; PRIMs.

5.1. Introduction

Children's development of reasoning about other people's representational mental states such as beliefs, desires and knowledge has been one of the most studied areas in developmental psychology. In order to conclude that an agent has such a theory of mind (ToM, Premack & Woodruff, 1978), Dennett (1978) argued that it is necessary to test whether the agent can correctly attribute a false belief to another agent. Since then, the explicit false belief task (Wimmer & Perner, 1983) has become one of the most commonly used tasks that verbally tests children's ToM. In the explicit first-order false belief task, children are required to make and report a decision about another person's mental state while they know the real situation, which happens to be different from the other person's false belief. Various studies have shown that children cannot pass the explicit first-order false belief tasks until the age of four (Wimmer & Perner, 1983; Wellman, Cross & Watson, 2001).

One of the most commonly studied explicit first-order false belief tasks is called the unexpected location change task. In this task the story goes more or less as follows: 'Sally and Anne are in the room. Sally puts her chocolate into the basket. After that, she leaves the room. Anne takes the chocolate from the basket and puts it into the box and she also leaves the room. Later, Sally comes back to the room.' The first-order false belief question is "Where will Sally look for the chocolate?" If a child correctly reasons about Sally's mental state, s/he reasons that because Sally did not see Anne taking the chocolate from the basket and putting into the box, Sally will look for the chocolate in the place where she last saw it—thus, the child would answer that Sally will look in the basket.

Interestingly, until the age of 4, children make systematic errors by reporting the real location of the chocolate, which is the box in the above story. This phenomenon is called 'reality bias' (Mitchell et al., 1996). Previous studies of the explicit false belief task showed that 3-year-old children's accuracy is around 30%, 4-year-olds' accuracy is around 50%, 6-year-olds' accuracy is around 80%, and finally around the age of 8, children's performance is at ceiling, similar to adults' performance (Wellman, Cross & Watson, 2001). According to the 'reality bias' view, in order to give correct answers, children should inhibit their own response and take into account others' perspectives.

What kind of cognitive skills are required for children to overcome their 'reality bias' and pass the explicit first-order false belief task? It is a matter of debate whether the development of first-order ToM is purely a matter of conceptual change. In fact, it has been shown that other cognitive factors contribute to the development of first-order false belief reasoning. Several studies have examined the so-called 'far transfer' of skills by training children with different cognitive

tasks and investigating whether children's performance on the first-order false belief task has improved or not after the training. Those studies revealed that there is indeed a far transfer of skills from language (Hale & Tager-Flusberg, 2003) and inhibition (Kloo & Perner, 2003) to first-order false belief reasoning. We believe that the working memory strategies that children use also contribute to the development of false belief reasoning. The important role of working memory for first-order false belief reasoning has already been shown by correlational studies (Gordon & Olson, 1998; Hughes, 1998; Keenan, Olson, & Marini, 1998). Moreover, we have evidence for a significant effect of the complex working memory task but not the simple working memory task in second-order false belief reasoning (Arslan, Hohenberger, & Verbrugge, 2017b). However, there has so far been no experimental training study focused on the role of working memory strategies in the development of first-order false belief reasoning.

Training studies need more time and effort than correlational studies. For this reason, constructing computational cognitive models to predict what kind of skills might be transferred to another domain (far transfer) is an effective way of designing an appropriate training study. There have been a few computational models of the development of explicit false belief reasoning (Wahl & Spada, 2000; Triona, Masnick & Morris, 2002; Bello & Cassimatis, 2006; Hiatt & Trafton, 2010; Arslan, Taatgen & Verbrugge, 2013; 2017a). However, none of those models are aimed to predict and explain far transfer from daily life tasks to explicit false belief reasoning.

In the current study, we aim to investigate the possible transfer of cognitive skills from working memory strategies that children use in their daily-life tasks to first-order false belief reasoning by constructing a computational cognitive model that helps us to make more precise predictions. To investigate the role of working memory strategies, we modeled one simple working memory task (the pencil task) and one complex working memory task (the marble task) together with the first-order false belief task. The pencil and marble tasks were inspired by Brain Quest game cards for children of ages 5 to 6 (<http://www.brainquest.com/>) and they differ from each other in terms of the complexity of the working memory strategies required to solve them (see the sections "A cognitive model of the pencil task" and "A cognitive model of the marble task" for details). We hypothesized that there would be more transfer from the marble task to the first-order false belief task than from the pencil task, because of the more complex working memory strategies required by the marble task, which are also necessary in the false belief task.

In order to model transfer from the pencil and the marble tasks to first-order false belief reasoning, we modeled the tasks using the cognitive architecture PRIMs (Taatgen, 2013). The PRIMs architecture implements the primitive elements theory (Taatgen, 2013) of the nature and transfer of cognitive skills. PRIMs

builds on the symbolic computational cognitive architecture Adaptive Control of Thought-Rational (ACT-R; Anderson, 2007) and it uses ACT-R modules, buffers and mechanisms such as production compilation (Taatgen, 2002).

The PRIMs architecture (Taatgen, 2013) breaks down the complex production rules typically used in ACT-R models into the smallest possible elements (PRIMs) that move, compare or copy information between modules. There is a fixed number of PRIMs in the PRIMs architecture. When PRIMs are used often over time, production compilation combines them to form more complex production rules. While those PRIMs may have some task-specific elements, PRIMs also have task-general elements that can be used by other tasks. Transfer occurs if two tasks have common task-general elements: One task can benefit from another trained task because of the already compiled production rules that are learned through production compilation. Taatgen (2013) showed the predictive power of Acttransfer by modeling a variety of transfer experiments such as text editing (Singley & Anderson, 1985), arithmetic (Elio, 1986), and cognitive control (Chein and Morrison, 2010).

In the following sections of this chapter, we will explain our PRIMs models in detail, present the results of the simulations and discuss our findings.

5.2. A cognitive model of the first-order false belief task

Our PRIMs model for the first-order false belief task was inspired by Arslan, Taatgen and Verbrugge's (2013; 2017a) ACT-R model and Wierda and Arslan's (2014) PRIMs model of first- and second-order false belief reasoning. A simulated storyteller presents the first-order false belief story to our model. The way we implemented this is by updating the perceptual buffer every 4 seconds with new story facts. For each picture in the story, the storyteller tells what happens in that particular picture. The model "listens" to the story and stores what happened in each picture in its declarative memory. The pictures that have actions related to changing the location of the object of interest are chained together in chronological order. Adding a pointer that refers to the previous picture fact realizes the chaining of the picture facts. Also, all related action facts are linked in a similar manner with the corresponding picture fact.

At the end of the story, the storyteller presents the model with a first-order false belief question ('Where will Sally look for the chocolate?'). First, the model creates a first-order chunk in declarative memory that represents the first-order false belief question ("Where will Sally look for the chocolate?"). Next, the model creates a zero-order chunk that represents the corresponding zero-order question ("Where is the chocolate?") by breaking up the first-order false belief

question. The model keeps a reference to the zero-order chunk in working memory, which in turn has a pointer towards the first-order chunk. After the question is presented, the model uses two strategies to reason about the question. The first strategy is a memory strategy in which the model always tries to retrieve a picture fact that has an action related to the object's location change. It then looks at that picture when remembering facts about it, such as "Anne put the chocolate into the box". The second strategy is a perception strategy, which is used whenever the model has forgotten the story facts. The model looks at each picture in detail and extracts the story facts from the picture. Below, we present the details of these two strategies (memory and perception) in detail.

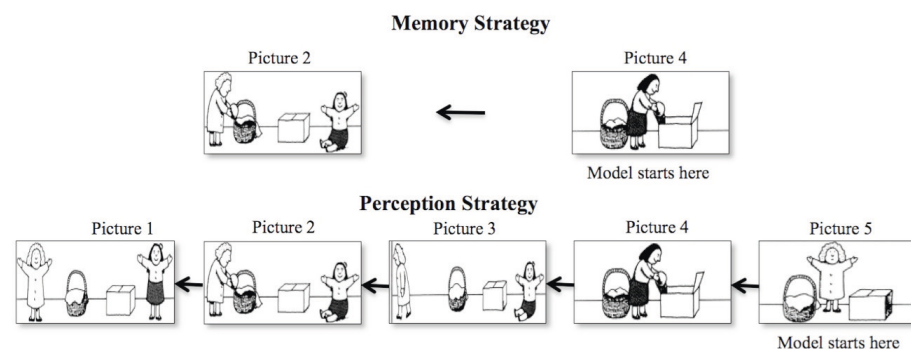


Figure 5.1. The order of the pictures that the false belief model attends in memory and perception strategies

5.2.1. The memory strategy

The memory strategy is the first strategy that the model uses. The model tries to retrieve what was the last picture in which an action happened that was related to the location of the object. If it retrieves that picture fact, it then tries to remember what exactly happened in that picture (for example, a location change of the chocolate). If the model successfully remembers that Anne put the chocolate into the box, it puts the location of the chocolate ("the box") in its working memory and then tries to recall the question. First, the zero-order question is retrieved by the reference that is kept in working memory. If the zero-order chunk does not point to a first-order chunk, the model gives an answer by reporting the location from its working memory ("the box"). However, in this particular task, the actual question put to the model is the first-order false belief question.

Thus, the model then tries to recall the first-order question ("Where will Sally look for the chocolate?"). If it retrieves the first-order question, it checks whether the person in the question performed the action in that picture. Because it was not Sally but Anne who put the chocolate into the box and Sally is absent in the picture, the model tries to retrieve another picture fact at which another action towards the object happened and again it tries to recall what exactly happened in that picture (Figure 5.1). This process continues until the person who moved the chocolate is the same person who is mentioned in the question.

If the model's run-time passes a preset threshold, the model stops reasoning and answers whatever it currently has in working memory. In this way, we simulate that the model gives up for whatever reason (for example, it takes too long or it gets distracted). As a result, the model will at first give either no answer at all or a zero-order answer. Note that this is because the model first stores the most recent location of the chocolate in its working memory, which corresponds to the zero-order answer ("the box"). When the model reaches the part of the story where the first-order answer ("the basket") can be found, this location will be stored in working memory and the model starts giving the correct first-order answer.

5.2.2. The perception strategy

In our behavioral study (Arslan, Verbrugge, Taatgen, & Hollebrandse, 2015a; Chapter 3 of this thesis), we have successfully trained 5- to 6-year-old children to pass the second-order false belief tasks. We experienced that on most occasions, children look back in the pictures. Similarly, our model uses the perception strategy by looking at the pictures in more detail if it fails to apply the memory strategy because it has forgotten some of the facts of the story as told by the storyteller. In the perception strategy, the model first focuses its attention at the most recently seen picture and inspects whether there is an action related to the salient object in the picture. If there is a person present in the picture, it checks whether this person performed an action or not. Subsequently, it creates a new action fact about the picture in memory and starts to reason with those newly created chunks in the same way as in the memory strategy.

Note that both the perception strategy and the memory strategy use almost the same mechanism for reasoning about the question. The difference is that the irrelevant pictures for finding the answer are skipped in the memory strategy, whereas every picture has to be inspected in the perception strategy (see Figure 5.1). This is because the memory strategy broke down, and the model cannot immediately recall Picture 4 and subsequently Picture 2 of the false belief story (see Figure 5.1) at which there are actions related to the location of the object.

5.3. A cognitive model of the pencil task

As we mentioned above, we modeled one simple working memory task, the pencil task, and one complex working memory task, the marble task. In the former task, the goal is to count the total number of yellow and green pencils in a group of blue, red, yellow and green pencils (Figure 5.2). We modeled this task as follows. The model first looks at a pencil that is in its perceptual buffer. If the color of the pencil is blue or red, it focuses its attention to another pencil. This procedure is repeated until the model finds a yellow or green pencil. It then initializes counting by retrieving a counting fact from its declarative memory and copying the retrieved number to the working memory. It keeps on searching pencils until it finds another yellow or green pencil. When it finds one of those, the counter in working memory is updated by retrieving the next counting fact. After attending all pencils, the model reports the total number of yellow and green pencils. As becomes clear from this explanation, this task does not need any complex working memory strategies. It simply uses one slot in the working memory buffer and it updates that slot whenever it is necessary.

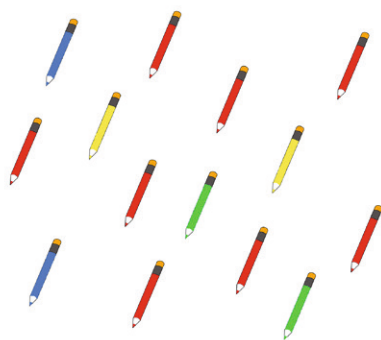


Figure 5.2. The pencil task (simple working memory)



Figure 5.3. The marble task (complex working memory)

5.4. A cognitive model of the marble task

The goal of this task is to find, out of a small number of bags of marbles, the two bags that contain the same number of marbles of the same color (Figure 5.3). Our model uses a strategy that focuses on one color in a bag and counts that color of marbles in each bag until finding a bag that shares the same number of that color. We assume that this is one of the strategies that children use in general. Because we are interested in comparing a complex working memory strategy with a simple one, the strategy that we used for modeling will suffice for our purposes.

The model starts by looking at the first bag and retrieving a color fact from its declarative memory to count the marbles of that color. For example, if the model retrieves the color red, it copies red to one of four working memory slots. At the same time it copies the identity of the bag (Bag-1) in another working-memory slot to report it back when necessary. Then, it looks at the first marble that is in its perceptual buffer, which is blue in our example. Since blue is not the same as the color that is in working memory (red), the model focuses its attention to another marble and repeats that procedure until it finds a marble that matches the color in working memory. After it finds a red marble, it initializes counting by requesting the retrieval of a counting fact from its declarative memory and copying the retrieved number to a third working memory slot. The model then updates that counting slot if it attends another red marble.

Once all marbles of the current color in the current bag (Bag-1) are counted, the model tries to remember if it has already seen another bag that has the same number of marbles of the same color. In the example, because it is the first bag, the model cannot remember a bag that has the same number of red marbles and focuses its attention on another bag to continue to count the red marbles. It carries out the same procedures for the second and the third bags.

After counting all the red marbles in all bags and not remembering any bags that have the same number of red marbles, the model creates a new working memory chunk by emptying all its slots except the slot that has the current color (red). This process also consolidates all information present in working memory and thus creates a new chunk in declarative memory that can be retrieved later on—effectively it remembers which bags it has seen with how many marbles of a given specific color.

Later, it repeats the procedures above by retrieving another color from its declarative memory. Let's say the color blue is retrieved this time. The model counts the blue marbles in the first and second bags, and checks if they have the same number of blue marbles. Because this is not the case, it moves its attention to the third bag and counts the blue marbles. At this point the model can successfully retrieve the first bag with the same number of blue marbles, which is 1, from its declarative memory. Finally, it gives an answer by reporting the first and third bag.

5.5. Results

In order to investigate transfer from the simple working memory task (pencil) and the complex working memory task (marble) to the first-order false belief task, we ran simulations in three conditions. In the first condition (FB-only), we ran 100 simulations of a child doing the first-order false belief task 100 times (thus, a total of $100 \times 100 = 10,000$ trials were simulated).

In the second condition (Marble-FB), we first ran the marble task for 10,080 minutes (24 hours in 7 days) in ACT-R's time. The model would perform as many trials as it could possibly do within that time. Subsequently, the model performed 100 trials of the first-order false belief task. This condition was also simulated 100 times, simulating 100 children.

In the third condition (Pencil-FB), we followed the same protocol as in the second condition but first we ran the pencil task instead of the marble task. Table 5.1 shows the mean and the standard deviations of the number of simulations for each task. As can be understood from Table 5.1, the model could squeeze more trials of the pencil task than trials of the marble tasks into the 10,080 minutes. After all, each trial of the marble task, in which several numbers of objects need to be compared, takes much more time than the pencil task, which just involves counting an easily recognizable subset of objects. Therefore, the model has much more previous experience as expressed in number of trials in the pencil task before we run the false belief task model compared to if it is first trained with the marble task. However, as mentioned above, the amount of exposure as expressed in seconds is equal for both tasks.

Table 5.1. The mean and the standard deviations of the number of simulations for each task

Task	Mean	Standard Deviation
FB-Only	100	0
Marble-FB	217	8.6
Pencil-FB	850	21.0

Figure 5.4 shows the results of the simulations. In the FB-only condition, in which the model starts without any prior knowledge other than the PRIMs as described in the Introduction, the first-order false belief task model gives the zero-order answer (“reality bias”) by reporting the real location of the chocolate (i.e., “the box”) until around the 60th trial. After that, it gives the correct answer for the first-order false belief question (i.e., “the basket”).

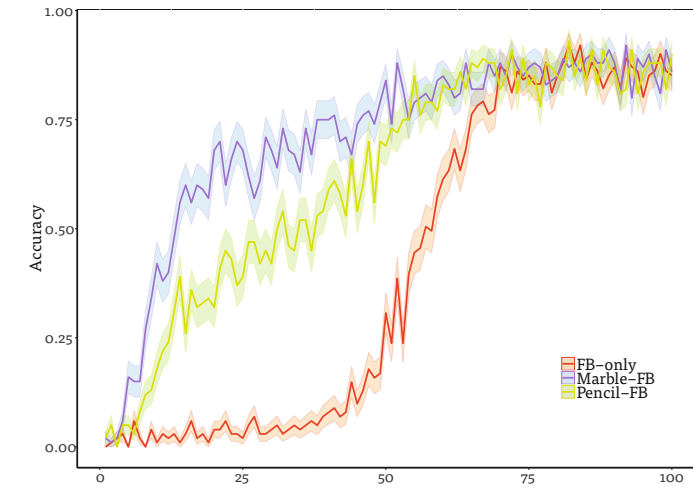


Figure 5.4. The results of the simulations of 100 trials averaged over 100 runs representing 100 children. The FB-only condition represents 100 trials of false belief task simulation only. The Marble-FB condition represents 100 trials of false belief task simulation after 10,080 minutes (ACT-R time) of training on the marble task. The Pencil-FB condition represents 100 trials of false belief task simulation after 10,080 minutes (ACT-R time) of training on the pencil task.

In the Marble-FB condition, in which the first-order false belief task model experienced the prior practice of the marble task, the model starts to give the correct answer much earlier, around the 15th trial. Finally, in the Pencil-FB condition, the model starts to give the correct answer for the first-order false belief question around the 35th trial, which is earlier than in the FB-only condition, but later than in the Marble-FB condition.

5.6. Discussion

Our goal was to investigate the role of working memory (WM) strategies in the development of first-order false belief reasoning. In order to achieve this goal, we modeled two real life examples, the pencil task and the marble task, corresponding to a simple and a complex working memory strategy, respectively, by using the cognitive architecture PRIMs.

In agreement with the previous behavioral studies that have shown the correlation between working memory and the development of first-order false belief reasoning (Gordon & Olson, 1998; Hughes, 1998; Keenan, Olson, & Marini, 1998;

see Arslan, Hohenberger & Verbrugge, 2017b for second-order false belief reasoning), our results show that having an experience with tasks that need working memory strategies contribute to this development. Because more complex working memory strategies are needed in our first-order false belief task model than a simple strategy that needs to just update the WM, we predicted that there would be more transfer from the marble task (complex working memory) to the first-order false belief task than from the pencil task (simple WM) to the first-order false belief task. The results confirm our hypothesis.

The first-order false belief task model learns to pass the task faster when it has a prior experience of a task that needs simple or complex WM strategies. This result is straightforward, as we compare the simulations with prior knowledge to a model that has no prior experience at all. More interestingly, the model that was first trained in the marble task, which required complex working memory strategies, mastered the first-order false-belief task much faster—even though the model was able to do fewer trials of the marble task in a given time period (M_{no} of simulations = 217, $SD = 8.6$) than the model that was first trained in the pencil task, which required simple WM strategies (M_{no} of simulations = 850, $SD = 21.0$). Note that the amount of exposure to both models was similar in terms of time, as stated above. Together with the experimental training studies that we mentioned in the Introduction, our work implies that passing false belief tasks is not a skill acquired through maturation, but by experience.

5.7. Future directions

Although the amount of exposure-time in the Marble-FB and the Pencil-FB conditions was the same, one could argue that it is the general complexity of the marble task (complex working memory), which causes the transfer to the false belief task. In addition to comparing the marble task to the pencil task (simple working memory), including a third task that has the same complexity as the marble task but that does not require complex working memory strategies might be a better control condition. Also, finding a task to model that has the same complexity as the first-order false-belief task but without the need of working memory might be worthwhile.

The results of our simulations suggest conceptual predictions that should be tested experimentally with 3- to 4-year old children.

Chapter 6: Cognitive Control Explains the Mutual Transfer Between Dimensional Change Card Sorting and First-order False Belief Understanding: A Computational Modeling Study on Transfer of Skills

*In which we investigate the mutual far transfer
between cognitive control and first-order false
belief reasoning by constructing computational
cognitive models using the cognitive
architecture PRIMs.*



This chapter was previously published as:

Arslan, B., Verbrugge, R., & Taatgen, N. (2017c). Cognitive control explains the mutual transfer between dimensional change card sorting and first-order false belief understanding: A computational modeling study on transfer of skills. *Journal of Biologically Inspired Cognitive Architectures*.
<http://dx.doi.org/10.1016/j.bica.2017.03.001>

Abstract

While most 3-year-olds fail both in the false belief task of theory of mind and Dimensional Change Card Sorting task of cognitive control, most 4-year-olds are able to pass these tasks. Different theories have been constructed to explain this co-development. To investigate the direction of the developmental relationship between false belief reasoning and cognitive control, Kloo and Perner (2003) trained 3-year-olds on the false belief task in one condition and on the Dimensional Change Card Sorting task in another condition. They found that there is a mutual transfer between the two tasks, meaning that training children with the Dimensional Change Card Sorting task with feedback significantly improved children's performance on the false belief task and vice versa. In this study, we aim to provide an explanation for the underlying mechanisms of this mutual transfer by constructing computational cognitive models. In contrast to the previous theories, our models show that the common element in the two tasks is two competing strategies, only one of which leads to a correct answer. Providing children with explicit feedback trains them to use a strategy of control instead of using a simpler reactive strategy. Therefore, we propose that children start to pass the false belief and cognitive control tasks once they learn to be flexible in their behavior depending on the current goal.

Keywords: false belief reasoning; cognitive control; transfer of skills; computational cognitive modeling; cognitive development; PRIMs

6.1. Introduction

There are many hilarious videos on the Internet showing 2- and 3-year-olds' failure on the *hide and seek* game and on the *marshmallow* test. On the other hand, most 4-year-olds are able to hide themselves at a place where the seeker cannot find them immediately in the *hide and seek* game. In the *marshmallow* test, most 4-year-olds are able to wait for the experimenter to come back to the room in order to get more marshmallows instead of eating one marshmallow right away. The key element of success in the *hide and seek* game is to be able to take the perspective of the seeker and the key element in the *marshmallow* test is to have self-control.

In line with these videos, a number of correlational studies have shown that there is a relation between children's development of *theory of mind* and *cognitive control* (Perner & Lang, 1999; Müller, Zelazo, Imrisek, 2005; Henning, Spinath, & Aschersleben, 2011). Theory of mind can be defined as a general term for perspective taking by reasoning about others' representational mental states such as beliefs, desires and knowledge (Premack & Woodruff, 1978). Cognitive control, which is an important component of executive functions, can be defined as the ability to flexibly select actions in the furtherance of chosen goals, instead of inflexibly reacting to the environment while ignoring the current goal. Therefore, cognitive control requires selecting appropriate information related to the current goal for processing and inhibiting inappropriate information and responses. For example, to succeed in the *marshmallow* test, children have to inhibit the urge to eat the marshmallows right away and have to consider the current goal, which is waiting for the experimenter in order to receive a larger award. Similarly, if an agent's initial goal is to find another agent who has blue eyes and if the current goal is finding an agent who has brown shoes, then the agent should ignore the eye color of other agents and attend to the agents' shoe color.

There are three main theories about the relation between theory of mind and cognitive control²¹. The *Cognitive Complexity and Control-revised* theory (CCC-r; Zelazo, Müller, Frye, & Marcovitch, 2003) suggests that the common component between theory of mind and cognitive control is representational and also related to the activation and inhibition of rules. According to this theory, theory of mind and cognitive control tasks develop together because they both require a child to reason by using embedded *if-if* rules and both need inhibition of rules. The second theory suggests that being able to take the perspective of others improves children's cognitive control abilities, meaning that there is transfer of skills from theory of mind to cognitive control (Perner, 1998). On the contrary,

²¹ see Carlson, Moses and Hix (1998), Leslie and Polizzi (1998), Carlson, Moses and Breton (2002) for other theories that are related to the role of other components of executive functions, such as inhibition and working memory.

the third theory suggests that the direction of transfer is from cognitive control to theory of mind (Russell, 1996 as cited in Kloo & Perner, 2003).

Although correlational studies have shown that children's theory of mind and cognitive control abilities co-develop, as reflected in the second and third theories, there is no consensus on the direction of this relationship. In order to investigate the direction of the relationship, Kloo and Perner (2003) conducted a training study with children by using a theory of mind task and a cognitive control task. We provide the details of these tasks in the following subsection. Kloo and Perner's results showed that there is a mutual transfer between cognitive control and theory of mind, meaning that training children with a cognitive control task with feedback significantly improved children's performance on a theory of mind task and vice versa.

Based on these findings, Kloo and Perner propose that the common component between the two tasks is representational. Differently from CCC-r theory, they argue that the problem 3-year-olds encounter is related to failure in re-describing an object or situation and that training children with explicit feedback helps them to understand that an object or certain situation can be described differently from different perspectives. However, Kloo and Perner stated that the exact nature of transfer effect remains to be determined.

The main goal of the current study is to provide an explanation for the nature of the mutual transfer between cognitive control and theory of mind by constructing computational cognitive models.

How does training children help transfer of skills? According to the *primitive information processing elements* theory (PRIMs; Taatgen, 2013), there are two explanations for the transfer of skills that can be modeled with the same mechanism. According to Explanation 1, skills can transfer from one task to another when those tasks have a substantial overlap in their procedural knowledge. For example, multi-column multiplication shares knowledge with multi-column addition, and many other pen-and-paper arithmetic algorithms. Acquiring this knowledge is a relatively slow process. On the other hand, Explanation 2 assumes that the knowledge for both tasks is already present in memory: it just has to be mobilized at the right moment. Suppose a particular task has two possible strategies, A and B, and suppose B is superior to A, but A is simpler. If parts of strategy B, in particular the parts that are necessary to select B, are trained in another task, it becomes more likely that strategy B will be chosen over strategy A. Our models are based on Explanation 2, because the training time in the experiment is relatively short.

In the following subsection, we first present the details of the theory of mind and cognitive control tasks that were used in Kloo and Perner's training study together with a summary of the design of the study, in order to provide a sufficient

background to understand our computational cognitive models and to interpret the simulation results.

6.1.1. Kloo and Perner's training study

Kloo and Perner's training study (Experiment 2) tested a sample of 44 children between the ages three and four ($M = 45.1$ months, $SD = 4.9$ months) at four different sessions almost one week apart from each other: i) pre-test, ii) training day 1, iii) training day 2, and iv) post-test.

At the pre-test and post-test sessions, children were tested with a standard theory of mind task and a cognitive control task together with a verbal intelligence task.

As a theory of mind task, Kloo and Perner used a standard false belief task (FB; Wimmer & Perner, 1983), which is one of the most commonly used tasks to assess young children's development of theory of mind. During the FB task, children listened to a story accompanied by illustrations showing that a protagonist placed an object into a location, after which that object was moved to another location while the protagonist was not present. Children had to *predict* where the protagonist would look for the object based on the protagonist's false belief, instead of reporting their own true belief about the location of the object. After that, children were shown another picture of the protagonist searching for the object based on her false belief (empty location) and were asked to *explain* the behavior of the protagonist. The same type of story with a different object and protagonists was used at the post-test session. On each false belief task, children's scores were between 0 and 2 based on their answers for the prediction and explanation questions, not their explanations. Children did not get any feedback at the pre-test and post-test sessions. Note that Kloo and Perner also reported children's performance on the predictions separately. For the purpose of our study, we only modeled children's predictions.

As a cognitive control task, they used the Dimensional Change Card Sorting task (DCCS; Frye, Zelazo, & Palfai, 1995). In the standard version of the DCCS task, children are presented with two target cards, one on the left and the other one on the right. After that, an experimenter introduces a set of test cards. The test cards have two dimensions, one of which matches with one target card and the other matches with the other target card (see Figure 6.1). At the beginning of the experiment (pre-switch phase), children are introduced to the rule of the "Animal" game. In the "Animal" game, children are expected to sort the test cards by pointing to the target that matches to the test card with the animal type. For example, if the test card "small horse" is shown, the children are expected to point to the

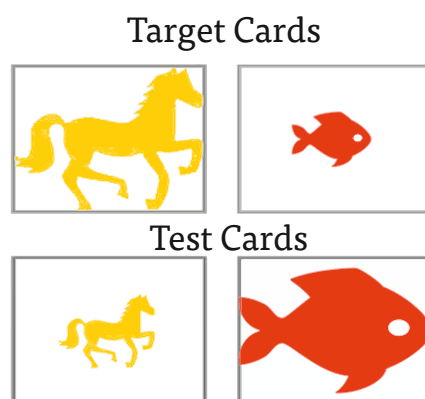


Figure 6.1. An example of the DCCS task. In this example, if the game is an “Animal” game, children are expected to sort the test card “small horse” by pointing to the target card “big horse”, which is on the left and to sort the test card “big fish” by pointing to the target card “small fish”, which is on the right. If the game is a “How-Big” game, children are expected to sort the test card “small horse” by pointing to the target card “small fish”, which is on the right and to sort the test card “big fish” by pointing to the target card “big horse”, which is on the left.

target card “big horse”, which is on the left. After playing six trials of the “Animal” game, the experimenter introduces the new “How-Big” game (post-switch phase). At the post-switch phase, children have to sort the test cards based on the other dimension, namely size, also for again six trials by pointing to the target card that matches to the test card with size. For example, if the test card “small horse” is shown, children are expected to point to the “small fish” target card, which is on the right. Even though most children around the age of three do not have major problems in sorting the cards correctly at the pre-switch phase, after the rule changes, they keep sorting the cards by the pre-switch rule instead of the new post-switch rule. On the other hand, similar to 4-year-olds’ development of false belief reasoning, most children around the age four are able to sort the cards correctly at the post-switch phase as well (Doebel & Zelazo, 2015).

Different from the pre-test, children were tested with a three-boxes version of the standard DCCS task at the post-test session, which had three target cards instead of two. The reason for using three target cards at the post-test session was to control for children’s usage of a reversal shift strategy, which is pointing out the opposite target card. Children were expected to sort six cards at the post-switch phase, both in the standard DCCS and three-boxes version of the DCCS tasks. The experimenter sorted the first card as an example; therefore, children’s score was between 0 and 5.

At the training sessions, children were assigned to one of the following three training groups: i) DCCS ($N = 14$), ii) FB ($N = 15$), iii) control ($N = 15$). Children in the DCCS group were trained with a DCCS task with three dimension switches (i.e., color, number, color, number). Subsequently, children were introduced with a new set of test cards while the target cards were the same and they were expected to sort the new test cards again first by color, then by number. Finally, a new set of test cards was introduced with new target cards and children were again expected to sort the cards first by color and then by number. Therefore, DCCS training consisted of ten switches in total for both training days. The experimenter provided positive and negative feedback by emphasizing which game they were playing and how they should sort the cards at each ten switches.

The crucial parts of the feedback for our model of DCCS at the training session are the parts in which children were reminded that they were not playing the pre-switch phase game anymore and were asked questions about which game they were playing at the post-switch phase (e.g., “... However, we are not playing the “Animal” game, the game with “horse” and “fish” (point), anymore. Now, we are playing the “How-Big” game. This is the game with “small” and “big” (point). What game are we playing now? Right/No. We are playing the “How-Big” game now. This is the game with “small” and “big” (point)...” We explain how this feedback helps children and how training on the DCCS task with this feedback transfers to improve children’s performance on the false belief task in the following section, “Modeling the Mutual Transfer between DCCS and False Belief Task (FB)”.

In the FB training group, children were trained with two false statements (Hale & Tager-Flusberg, 2003) and one FB task (Wimmer & Perner, 1983) at each training session. After each trial, the experimenter provided positive and negative feedback about their answers. Therefore, children got feedback six times in total at both training days. Similar to the DCCS training group, the feedback emphasized that the question was about the protagonist’s perspective, which was different from the children’s own perspective (e.g., “...Where is the shell now? Who put it there? Was Ernie able to see this?... Right/No. Ernie did not see that. So, does Ernie really know that the shell is in the red house? Right/No. Ernie does not know that the shell is in the red house now. Where does Ernie think the shell is? Right/No. Ernie still thinks that the shell is in the yellow tower...”).

In the control group, children were trained either with four relative clauses (Penner, 1999) or with three trials of a classic number conservation task (Piaget, 1965). Again, children got positive or negative feedback after each trial.

The results showed that there was a *transfer effect* from the DCCS task to the FB task, meaning that training children with the DCCS task by providing feedback significantly improved children’s performance on the FB task at the post-test

session. Similarly, there was a *transfer effect* from the FB task to the DCCS task, meaning that training children with the FB task by providing feedback significantly improved children's performance on the DCCS task at the post-test session. Moreover, there was a *training effect* of the DCCS task, meaning that training children on the DCCS task by providing feedback significantly improved children's performance on the DCCS task at the post-test session. Importantly, these improvements were significantly greater than children's improvement in the control group. Finally, although children's performance on the false belief task improved in all conditions, there was only a significant improvement in the DCCS training group. Kloo and Perner argued that the insignificant improvement of the FB prediction score in the FB training group might be due to the fact that children's scores were already good and there was little room for further improvements.

6.2. Modeling the mutual transfer between the DCCS task and the false belief task (FB)

In this subsection, we first discuss the relevant mechanisms of the cognitive architecture PRIMs and explain our DCCS and false belief task models. Subsequently, we explain the underlying mechanism of *training effect* in both training groups and the underlying mechanism of the *transfer effect* from the DCCS task to the FB task and vice versa. After that, we present the results of our simulations by comparing them to the experimental data from Kloo and Perner's (2003) training study. Finally, we introduce our models' predictions.

6.2.1. The relevant mechanisms of the cognitive architecture PRIMs

The cognitive architecture PRIMs is built as a theory of skill acquisition and of transfer of skills. It adopts the mechanisms of the declarative memory of ACT-R, which is a hybrid symbolic/sub-symbolic production-based cognitive architecture (Anderson, 2007).

Similar to ACT-R, the factual knowledge is represented in the form of chunks in declarative memory (i.e., "The president of the USA is Barack Obama"). However, in addition to the chunks of factual information, PRIMs architecture has *operators* (represented as hexagons in Figure 6.3) and *goals* (represented as rounded rectangles in Figure 6.3) in declarative memory. Operators, like production rules in ACT-R, are in the form of IF-THEN rules (condition-action) and implement the instruction of the given task.

The PRIMs architecture breaks down the complex production rules of ACT-R, which represent procedural knowledge (i.e., how to drive a car), into a fixed number of smallest possible elements, named PRIMs. PRIMs only move, compare or copy information between modules (i.e., declarative, visual, motor modules) independent from the content of the information. For example, a *condition* PRIM checks if working memory is empty and an *action* PRIM copies the visual input to working memory independently from the content of the information. Operators combine these PRIMs together to perform a task. Figure 1.6 presents the global outline of the PRIMs architecture.

For instance, in Figure 1.6, the operators, which are represented by the colored nodes, represent the task-specific operators of the DCCS and FB models and combine the gray (condition) and white (action) nodes, which represent the task-general *condition-action* PRIMs. While the red colored nodes denote the operators of the DCCS model, the blue colored nodes denote the operators of the FB model. The yellow halos show the common PRIMs between the two tasks.

In the PRIMs architecture, a single task is implemented by multiple goals that can be reused for other tasks. Unlike ACT-R's production rules, there is no hard connection between goals and operators in PRIMs (represented as dashed arrows in Figure 6.3), meaning that if a goal is triggered in a situation in which there are no associated operators, any matching operator can be tried. Current goals of the model activate operators to achieve those goals. If an operator is successful to complete a goal, the strength of association between the goal and the operator increases.

As we mentioned in the Introduction, the PRIMs architecture has two explanations that explain transfer of skills. Explanation 1 is based on the transfer of the task-general sequences of PRIMs. When a particular sequence of PRIMs is used often over time, it becomes more efficient to carry out that sequence. Whereas initially every PRIM is carried out individually, after learning the whole sequence of PRIMs is carried out in a single step (i.e., production compilation), considerably speeding up the process. Sequences of PRIMs are always task-general and can, therefore, be reused in other tasks. This means that if two tasks have common structural overlap, the PRIMs architecture can model knowledge transfer from one task to another. However, Explanation 1 is based on a slow compilation process and therefore transfer occurs relatively slowly.

Explanation 2 is based on training a particular strategy, which is represented by operators. Operators, like other chunks in declarative memory, have base-level activations and associative strengths. After training a model with a task that forces the model to use a particular strategy (e.g., a proactive strategy), when the model is presented with another task that has two competing strategies (e.g., reactive vs. proactive), the model chooses the trained strategy (e.g., proactive)

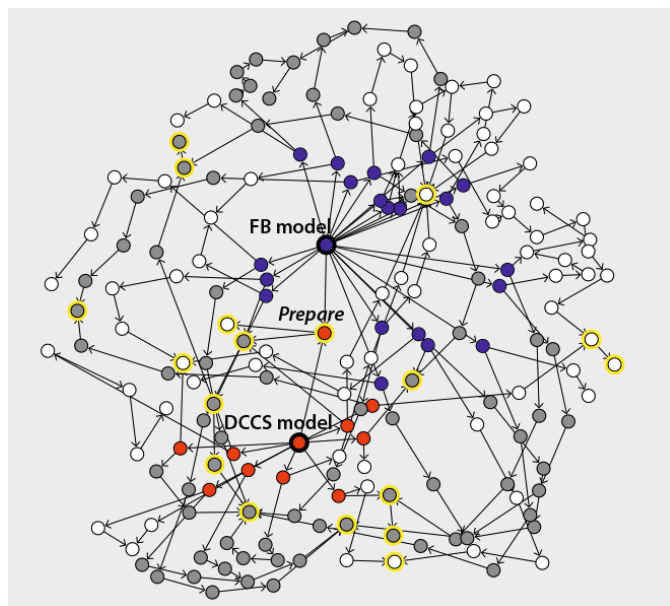


Figure 6.2. The representation of the operators (colored nodes) and condition-action PRIMs (gray and white nodes) in declarative memory for the FB model and the DCCS model. The yellow halos show the common PRIMs between the two tasks.

instead of the alternative competing strategy (e.g., a reactive strategy). Because Explanation 2 is based on the activations of the operators in declarative memory, transfer occurs faster than according to Explanation 1, which is based on utilities of PRIMs.

As can be seen from Figure 6.2, there is not so much overlap of condition-action PRIMs (gray and white nodes) between the DCCS and FB models and there is only one operator (i.e., *prepare*) that both models share. We argue that the key element of transfer from the DCCS task to the FB task and vice versa is based on Explanation 2 of the PRIMs architecture, which is training to choose a particular strategy, because the training time in the experiment is relatively short.

6.2.2. A model of the Dimensional Change Card Sorting task (DCCS)

We constructed a model of the DCCS task without and with feedback representing Kloo and Perner's pre-test/post-test sessions and the training sessions, respectively. The DCCS model with feedback has an additional goal and an operator

related to that goal that forces the model to prepare to use the *strategy of control* (see Figure 6.3a²²). We explain how the model uses the *strategy of control* in detail below.

In line with Kloo and Perner's experiment, the DCCS model without feedback at pre-test and post-test sessions first plays six trials of the "Animal" game at the pre-switch phase and, after that, plays six trials of the "How-Big" game at the post-switch phase. Again, in line with Kloo and Perner's experiment, the DCCS model with feedback at the training sessions sorts the cards with five switches.

The steps that the DCCS model goes through over time as follows (cf., Buss & Spencer, 2008; Morton & Munakata, 2002; Marcovitch & Zelazo, 2000; van Bers, Visser, van Schijndel, Mandell, & Raijmakers, 2011):

1. The model starts with the goal "store-game" which spreads activation to the operators that put which game the model is playing into working memory (e.g., "Animal", "How-Big") and sends this to the declarative memory to be retrieved later if it is necessary (Figure 6.3a, O1 and O2).
2. Subsequently, there are two competing strategies to choose from after a test card is presented and before attending to a dimension of a presented test card. If the *default-attend* strategy is selected, meaning that the *default-attend* goal has a higher activation in the declarative memory, the model attends a dimension of the test card based on the pre-switch phase (e.g., the type of the animal) without checking what the game was (Figure 6.3a, O3). Therefore, while this strategy leads the model to a correct answer in the pre-switch phase, it does not work for the post-switch phase because the goal of the post-switch phase is to attend to the size of the animals instead of the animal type.

On the contrary, if the model selects the *prepare* strategy, it prepares itself to use the *strategy of control*, meaning that it changes the current goal to *control* (Figure 6.3a). Unlike the *default-attend* strategy, the *strategy of control* first requests a retrieval of the current game (i.e., "Animal" or "How-Big"), which was stored in the declarative memory at the beginning of the task (Figure 6.3a, O5). In this way, the model uses cognitive control by being flexible in behavior based on the current goal. After that, the DCCS model focuses its attention on a dimension based on the retrieved game (Figure 6.3a, O6 or O7). Therefore, when the model uses the *strategy of control*, it gives correct answers most of the time both at the pre-switch and post-switch phases. For example, if the game is a "How-Big" game at

²² Note that the two models already have both a default strategy, which lacks of control, as well as a control strategy. This choice is based on the fact that Kloo and Perner's experimental results have shown that some children can pass these tasks even when they are presented to them for the first time. Our models are designed to represent an average child performing the tasks.

the post-switch phase and the test card is “small yellow horse”, the *default-attend* strategy focuses on “horse”, which is based on the pre-switch rule, “Animal” game. On the other hand, the *strategy of control* first retrieves what the game was (“How-Big”) and based on this retrieval, it focuses on the “How-Big” dimension of the test card, namely “small”.

- After focusing on a dimension of the test card, the DCCS model makes a decision by requesting a retrieval of one of the decision chunks (i.e., “big yellow horse” on the left and “small red fish” on the right) from its declarative memory (Figure 6.3a, O8). For instance, at the post-switch phase, while the *default-attend* strategy (“attend-animal”) focuses its attention to the dimension “horse” and gives the wrong answer “left” after retrieving the decision chunk “horse left”, the *strategy of control* (“attend-howbig”) focuses its attention to the dimension “small” and gives the correct answer “right” after retrieving the decision chunk “small right”.

There is an additional mechanism of the DCCS model that leads the model to make errors when the retrieval of a decision is requested. It has been shown that there is a visual clash between target and test cards (Doebel & Zelazo, 2015; Perner & Lang, 2002). For example, there is a visual clash between the picture of the big yellow horse on the target card and the small yellow horse on the test card when it needs to be sorted by size at the post-switch phase, which is after sorting the cards by animal type (pre-switch phase). In our DCCS model, this visual clash is represented by the strength of associations of chunks (S_{ji}) in declarative memory. As a result of the visual clash, although the model selects the correct *prepare* strategy that prepares the model to use the *strategy of control*, it can still make errors during retrieval.

Figure 6.4b shows an example of the associations between the “Animal” and “How-Big” types of chunks in the DCCS model. In addition to the positive associations with the same subgroup type of chunks (e.g., “horse – horse”), there are also positive associations between the subgroup of the “Animal” type of chunks and the subgroup of the “How-Big” type of chunks due to the visual clash (e.g., “horse – small”; “fish – big”). While the former positive associations lead the model to give correct answers when the correct strategy is selected, the latter positive associations represent the visual clash from the target cards and lead the model to make errors even if the correct strategy is selected. For example, at the post-switch phase, if the test card is “small yellow horse” and there is a target card on the left “big yellow horse” and on the right “small red fish”, even when the model uses the correct strategy (i.e., *strategy of control*) and focuses its attention on “small” according to the “How-Big” game, it can still give the wrong answer “left” instead of “right” based on the positive association between “small” and “horse”.

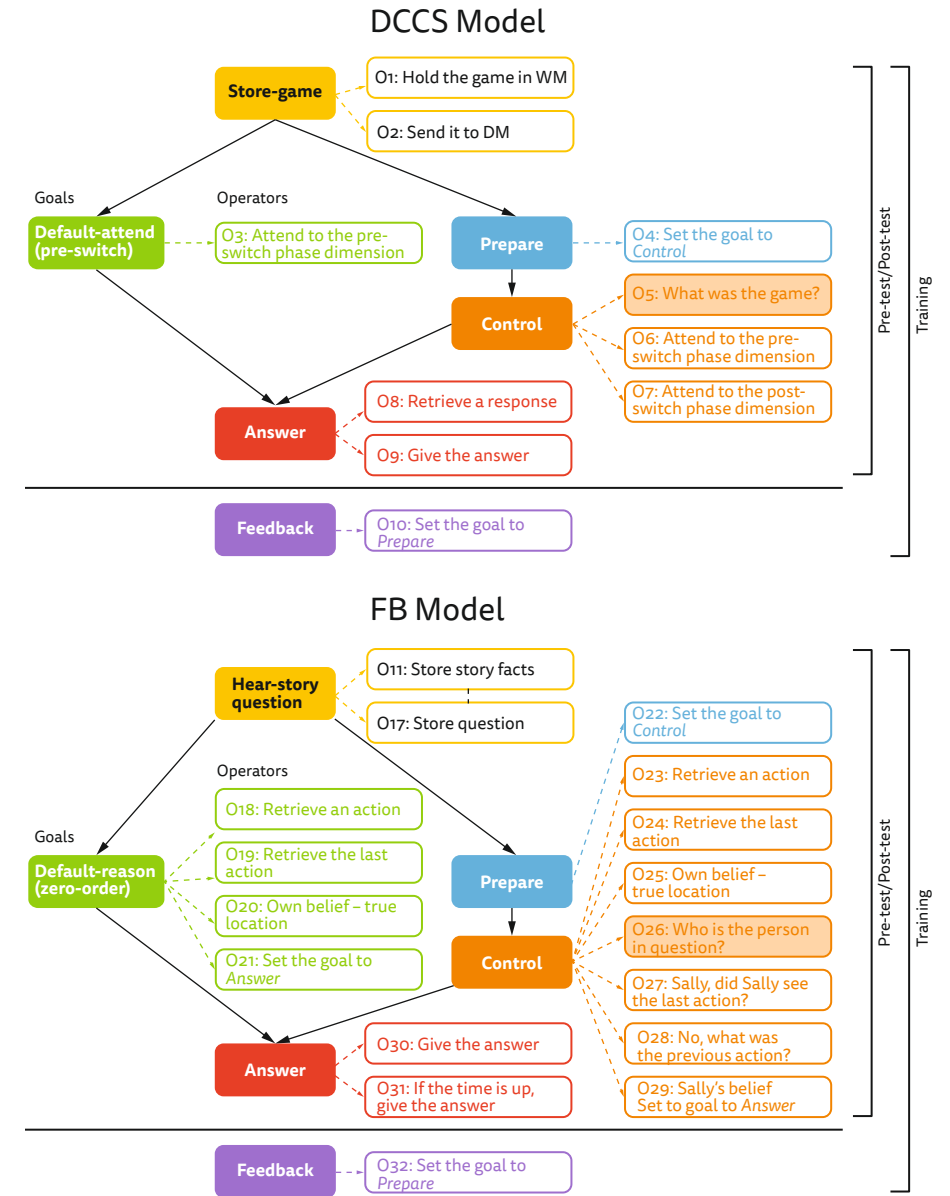


Figure 6.3. a) The DCCS model and b) the FB model at pre-test/post-test and training sessions. Note that bifurcations represent competing strategies²³ and the dashed arrows represent the operators related to the goals.

²³ Considering that children around the age of four start to use a strategy of control and that the tasks used in this study are novel tasks for children, we constructed our models with two different strategies (i.e., default and control). However, see Cohen, Servan-Schreiber, & McClelland (1992) for a framework proposing graded degrees of control.

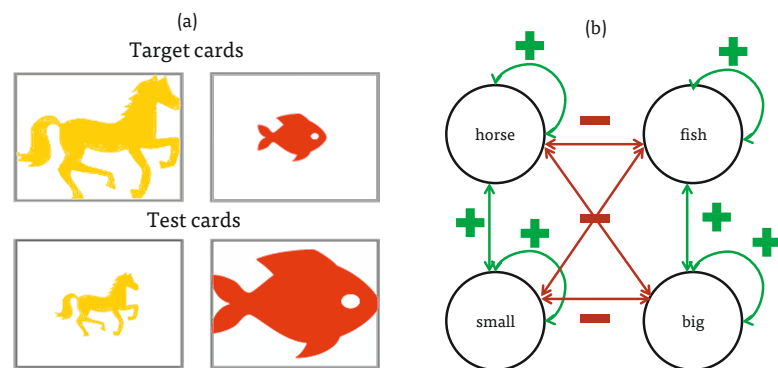


Figure 6.4. a) An example of the target and test cards in the DCCS task and b) an example of the associations between “Animal” and “How-Big” types of chunks in the DCCS model. The strengths of associations for the dimensions are set as follows: (horse horse 1.5), (fish fish 1.5), (small small 1.5), (big big 1.5), (big small -1.5), (horse fish -1.5), (horse big -1.0), (horse small 1.0), (fish small -1.0), (fish big 1.0).

6.2.3. A model of the false belief task (FB)

Similar to the DCCS model, we constructed a model of the FB task without and with feedback representing Kloo and Perner’s pre-test/post-test sessions and the training sessions, respectively. The FB model with feedback has an additional operator, which is associated to the goal “Feedback”. This additional operator forces the model to prepare to use the *strategy of control* whenever the model presented with feedback, meaning that it changes the current goal to *control* goal (see Figure 6.3b).

The steps that the FB model goes through over time as follows (cf., Bello & Cassimatis, 2006; Goodman et al., 2006; Hiatt & Trafton, 2010):

1. First, the story facts that include actions (e.g., “Ernie put the shell in the yellow tower”, “The bear put the shell in the red house”) and the false belief question (i.e., “Where does Ernie think the shell is?”) are presented on the screen one by one. The operators that are associated with the “Hear-story-questions” (Figure 6.3b, O11 – O17) put those facts into working memory and send them to declarative memory by chaining them together to be retrieved later when necessary. After being presented with the FB question, the model starts reasoning.
2. Similar to the DCCS model, the FB model has two competing strategies to choose from before starting to reason about the presented false belief question (i.e., “Where does Ernie think the shell is?”). The *default-reason*

strategy gives an answer based on the model’s own perspective (reality/zero-order reasoning strategy) without checking who is the question in person (i.e., as if the question was “Where is the shell” instead of the false belief question “Where does Ernie think the shell is?”). This strategy requests a retrieval of an action²⁴ (Figure 6.3b, O18). If the retrieved action is not the last action, an operator requests the retrieval of a last action (Figure 6.3b, O19). When the last action is retrieved (“The bear put the shell in the red house”), the model creates a “belief” chunk in working memory (Figure 6.3b, O20) about the location of the object (“in the red house”) and gives an answer based on its own belief (Figure 6.3b, O30).

Alternatively, if the model selects the *prepare* strategy, it prepares the model using the *strategy of control* by changing its goal to “Control” (Figure 6.3b, O22). The *strategy of control* starts with reasoning from the model’s own perspective as in the *default-reason* strategy (Figure 6.3b, O23 – O25). However, subsequently, it requests a retrieval about the person in question (“Ernie”) instead of giving an answer based on its own perspective (Figure 6.3b, O26). Note that this procedure is very similar to the DCCS models’ *strategy of control*, which first checks what the game was instead of the *default-attend* strategy that does not have an element of control (Figure 6.3a, O5).

3. After retrieving that the question is about “Ernie”, the FB model requests a retrieval whether “Ernie” saw the shell in the location that is in its working memory (“in the red house”, Figure 6.3b, O27). This retrieval request leads to a retrieval error. Based on this retrieval error, the model “infers” that “Ernie does not know that the shell is in the red house” and requests a retrieval of a chunk that includes “Ernie” and an action (Figure 6.3b, O28). Finally, the model retrieves the chunk “Ernie put the shell in the yellow tower” and creates a “belief” chunk in its working memory that “Ernie believes that the shell is in the yellow tower” (Figure 6.3b, O29) and gives the correct answer “in the yellow tower” (Figure 6.3b, O30).

In addition to selecting the wrong *default-reason* strategy, the FB model has another mechanism that leads the model to make errors. This mechanism is due to the time threshold of the FB model (i.e., 28 seconds). If the model’s run-time passes the preset threshold, the model stops reasoning and gives the location that it currently has in working memory as an answer (Figure 6.3b, O31). In this way, we simulate that the model gives up reasoning for any reason (e.g., it takes too long or it gets distracted). The idea of a time threshold when children are performing a task is consistent with research showing that children perform

24 We used the action of moving the shell, but the model could also easily be adapted for seeing.

better in language comprehension tasks and cognitive tasks when they are given more time (Ling, Wong, & Diamond, 2015; van Rij, van Rijn, & Hendriks, 2010; Hendriks, van Rijn, & Valkanier, 2007; Diamond, Kirkham, & Amso, 2002). When the model uses the *strategy of control*, first it starts to reason from its own perspective and puts into working memory the location where the shell really is (reality). If the time threshold is reached and if the model has not reasoned about Ernie's perspective yet, it gives the answer "in the red house" instead of the answer "in the yellow tower".

6.2.4. The underlying mechanism of the training effects

As we mentioned before, we use the term *training effect* to refer both to the improvement of the DCCS task after training on the DCCS task with feedback and to the improvement of the FB task after training on the FB task with feedback.

Similar to Kloo and Perner's experiment, both the DCCS and FB models receive feedback after each trial at the training sessions. Note that in the DCCS training group, the feedback forces children to first check which game they are playing before making a decision (i.e., "What game are we playing now?"). In the FB training group, the feedback urges children to take the perspective of the protagonist who is in question (e.g., "...Does Ernie really know that the shell is in the red house? Where does Ernie think the shell is? ..."). Therefore, the feedback forces children to use a *strategy of control* in the both DCCS and FB training groups.

The feedback in training sessions is represented as follows for both the DCCS and FB models: i) The screen that the model "sees" presents the word "feedback"; ii) the operator "feedback" (Figure 6.3a, O10; Figure 6.3b, O32) which is associated to the goal "Feedback" matches the current state of the model and puts the goal "Prepare" into one of the goal slots. With repetition, this procedure increases the activation of the *prepare* strategy, which forces the model to use the *strategy of control*. Therefore, the DCCS model starts to use a *strategy of control* by first checking what the game is instead of the *default-attend* strategy and the FB model starts to use the *strategy of control* by taking the perspective of the person in question instead of giving an answer based on its own perspective.

In this way, in the DCCS training group, after training the DCCS model with feedback, the accuracy of the DCCS model at post-test becomes higher than the DCCS model at pre-test. Similarly, in the FB training group, after training the FB model with feedback, the accuracy of the FB model at post-test becomes higher than the FB model at pre-test.

6.2.5. The underlying mechanism of mutual transfer between the DCCS and FB tasks

In the previous three subsections, we explained how the FB and DCCS models work and we delineated the underlying mechanisms of the training effect. In this subsection, we explain the underlying mechanism by which our models show transfer from the DCCS task to the FB task and vice versa.

As we mentioned before, there is not so much overlap in the procedural knowledge between the FB and DCCS models (Figure 6.2, the yellow halos). The key element of the mutual transfer between the DCCS and the FB models is based on the PRIMs architecture's Explanation 2, which is training a particular strategy.

As shown in Figure 6.3a and Figure 6.3b, the FB and DCCS models have a common structure. There are two competing strategies, only one of which leads the model to give a correct answer. While the *default-attend* and *default-reason* strategies lack the element of control, the *prepare* strategy, which is trained by the explicit feedback, forces the model use the *strategy of control* (see Taatgen, 2013 for a similar modeling approach to the transfer between Stroop task and task switching). Thus, once the DCCS model has been trained on the FB model with feedback, the activation of the *prepare* strategy gets higher. Because the *prepare* strategy's activation increases, at the post-test session, the DCCS model selects the *prepare* strategy instead of the competing *default-attend* strategy after training on the FB model with feedback. Therefore, the DCCS model's accuracy gets higher at the post-test session when it has been trained on the FB model with feedback. Similarly, once the FB model has been trained on the DCCS model with feedback, again the activation of the *prepare* strategy gets higher. Therefore, the FB model selects the *prepare* strategy instead of the competing *default-reason* strategy and the FB model's accuracy gets higher at the post-test session.

6.3. Results of the DCCS and FB models and comparison to experimental data

Similar to Kloo and Perner's study, we ran simulations in three training groups (DCCS training, FB training, control). Table 6.1 shows the protocol to obtain the results of the DCCS and FB models for each training group at pre-test, training and post-test sessions. We repeated the protocol in Table 6.1 for 15 times at each training group. In this way, we aimed the results to represent 15 children performing the tasks at each training group. For example, the accuracies of the DCCS and FB models at each pre-test and post-test sessions are based on a total of 1,500 repetitions (15 * 100) in each training group.

Table 6.1. The number of repetitions of the DCCS and FB models at pre-test, training and post-test sessions in three training groups.

Sessions	DCCS Training Group		FB Training Group		Control Group	
Tasks	DCCS Task	FB Task	DCCS Task	FB Task	DCCS Task	FB Task
Pre-test	DCCS * 100	FB * 100	DCCS * 100	FB * 100	DCCS * 100	FB * 100
	(reset)	(reset)	(reset)	(reset)	(reset)	(reset)
Training	DCCS with feed- back * 30	DCCS with feed- back * 30	FB with feed- back * 40	FB with feed- back * 40	DCCS * 30	FB * 40
Post-test	DCCS * 100	FB * 100	DCCS * 100	FB * 100	DCCS * 100	FB * 100

As can be seen from Figure 6.5a and Figure 6.5b, our results have similar patterns with Kloo and Perner's results in terms of training effect and transfer effect. Importantly, similar to Kloo and Perner's findings, the improvements of both the DCCS and FB tasks in the DCCS training group is higher than their improvements in the FB training group. Based on our results, we predict that the reason of this difference is due to the unequal number of times of feedback that was provided in the DCCS and FB training groups. Note that Kloo and Perner's experiment trained children ten times with feedback in the DCCS training group. On the other hand, children were trained six times with feedback in the FB training group. We simulated this by running 40 repetitions at the training session of the FB training group (40 times feedback) and 30 repetitions at the training session of the DCCS training group (30 * 5 times feedback). The higher number of repetitions in the DCCS training group makes the selection of the *strategy of control* more likely than the *default* strategies due to increased activation. Note that the only parameter that is fitted to the experimental results is the number of repetitions of the FB training group. If the number of repetitions of the FB training group is set proportional to Kloo and Perner's experimental design (i.e., 90), the improvement of the models in this training group will be higher than children's improvement in the experiment.

In order to predict what the results would be if both the DCCS and FB training groups received feedback the same number of times, we followed the same protocol as in Table 6.1, but we trained the DCCS and FB models with the same number of times feedback. We kept the number of repetitions of the DCCS model with feedback in the DCCS training group the same (5 * 30 repetitions). However, this time, we ran 150 repetitions of the FB model with feedback in the FB training group instead of 40 repetitions. The results showed that, similar to the DCCS

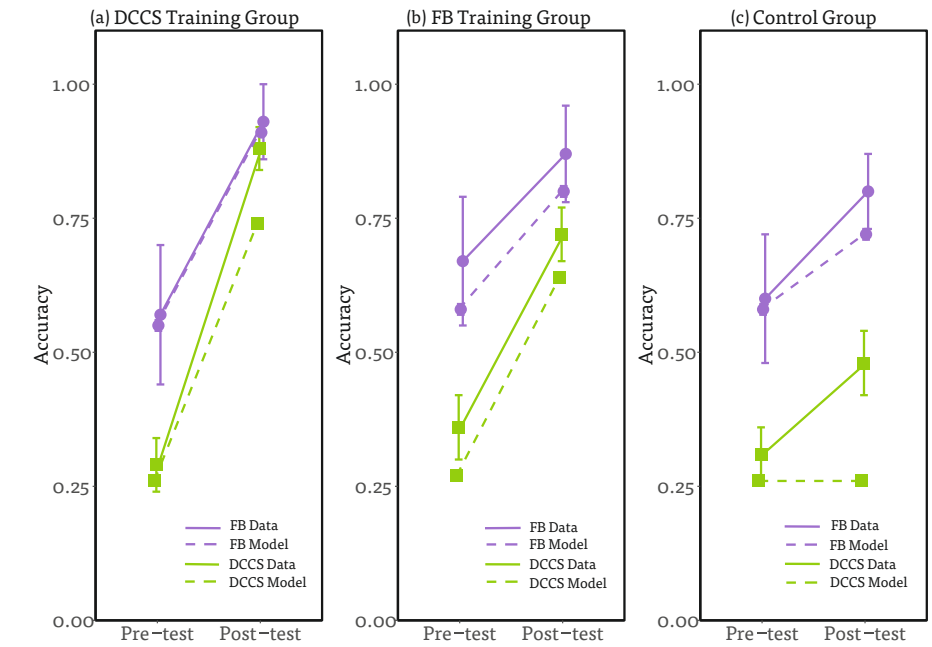


Figure 6.5.²⁵ The comparison of experimental data and the model results of the DCCS and the false belief (FB) task for a) DCCS training group, b) FB training group, and c) control group.

training group, the accuracy of the DCCS model would increase from 27% to 73% and the accuracy of the FB model would increase from 55% to 99% from pre-test to post-test sessions.

6.3.1. Effect of parameters

All the parameters were set to their default values, except the *retrieval threshold*, *input activation*, *default-operator-association*, *default-inter-operator-association* and *goal-operator-learning* parameters.

The *retrieval threshold* was set to -3 (default -2), in order to make sure that the errors are not due to just forgetting where the target cards were in the DCCS task (because in the experiment they were presented on the table) or to forgetting the story facts in the FB task. Thus, the DCCS model's failure is due to inappropriate strategy selection and the FB model's failure is due to the time threshold (28

²⁵ Because Kloo and Perner's study did not report the standard errors, the standard error bars of the FB and DCCS data were calculated based on the proportions under the assumption that there was no missing data.

seconds) and inappropriate strategy selection. The *input activation* parameter was set to 1 (default 0) in order to make sure that there is enough spreading activation from the input to the chunks in declarative memory. Moreover, we set the *default activation* parameter to 1 (no default) in order to specify the lower bound of base-level activation for all chunks in the models.

In line with the previous models of PRIMs (Taatgen, 2013), we set the *default-operator-association* parameter to 8 (default 4) in order to guarantee that operators relevant to one of the goals in the goal buffer are more likely to be retrieved instead of other operators. Also, to ensure that an operator that is associated to the same goal as the previous operator is selected, we set the *default-inter-operator-association* parameter to 2 (default 1).

Finally, we turned the *goal-operator-learning* parameter on. By turning on this parameter, the model learns the associations between the goal and retrieved operators by using reinforcement learning. The *reward* parameter (no default) that sets the maximum time to reach the goal, which is used in reinforcement learning, was set to 10 in the DCCS model and was set to 60 in the FB model. The differences in the reward parameters between the two models are due to the time needed for the models to complete the tasks. Note that all of the above-mentioned non-default parameters were set before running the simulations.

6.4. Predictions

1. In Kloo and Perner's study, the reason for the lesser improvements of the DCCS and FB tasks in the FB training group than in the DCCS training group is that children received fewer times feedback (6 and 10 respectively). If children would receive feedback the same number of times, the improvements in the FB training group would be similar to the improvements in the DCCS training group.
2. In the DCCS task, even when children use the *strategy of control*, it might be possible to make errors due to visual clash from the target cards. Similarly, even when children do not use the *strategy of control* at the post-switch phase, they can still give a correct answer based on the visual clash.

6.5. Discussion and Conclusions

In order to provide a procedural account for Kloo and Perner's (2003) findings that there is a mutual transfer between the Dimensional Change Card Sorting task (DCCS) and the false belief task (FB), we constructed computational cognitive

models of the DCCS and the FB tasks by using the PRIMs cognitive architecture. The main finding in this study is that using the *strategy of control* instead of a simpler strategy is the key factor for the transfer from the DCCS task to the FB task and vice versa. Based on our results, we argue that explicit feedback in the training groups trained children to use a *strategy of control*.

Our modeling approach explains the mutual transfer between the DCCS task and the FB task based on the PRIMs theory's Explanation 2, because the training time in the experiment is relatively short. While Explanation 1 is a slow process and explains transfer of skills if the tasks have common procedural knowledge, Explanation 2 is a fast process and is based on training a particular strategy. In both of the DCCS and FB models, there are two alternative strategies that compete with each other, only one of which leads the model to give a correct answer.

Explicit feedback trains both the DCCS and FB models to *prepare* to use a *strategy of control*, instead of giving an answer based on a simpler strategy. The *strategy of control* checks what the game was in the DCCS task and who was the person in question in the FB task before giving an answer. On the other hand, more simple strategies (i.e., *default-attend*, *default-reason*), which lack the element of control, focus on the pre-switch rule dimension without checking what the question was about in the DCCS task and reporting the model's own perspective in the FB task.

Kloo and Perner argue that explicit feedback in the DCCS and FB training groups help children to understand that an object or a certain situation can be described differently from different perspectives. Although we agree that children should be able to redescribe objects or situations from different perspectives in order to give correct answers, our computational modeling approach shows that the redescription hypothesis alone is not enough to explain how children learn to overcome the cognitive control elements of the DCCS and FB tasks.

Moreover, our models show that children do not necessarily need to represent *if-if* embedded rules (e.g., in Figure 6.1, *if the game is an "Animal" game then if the test card is a 'small yellow horse' then this card goes to the left box*) as the CCC-r theory suggests (Zelazo et al., 2003). Both the DCCS and the FB models' performance is based on requesting a retrieval from declarative memory and making decisions based on the retrieval. Therefore, they do not need to represent *if-if* embedded rules but they should both have a *strategy of control* to request the retrieval before making decisions.

Based on our models, we made two predictions. The first prediction, proposing that the lesser improvements of the DCCS and FB tasks in the FB training group than in the DCCS training group is due to the fact that children received fewer times feedback, should be tested by replicating Kloo and Perner's study by training children with the same number of times feedback in the DCCS and FB training groups. The second prediction, indicating that children can give correct

answers even when they apply the wrong strategy instead of the *strategy of control*, can be tested by looking at children's reaction times at the post-switch phase. Because using the *strategy of control* requires to request a retrieval from declarative memory and the decision is made based on the retrieval, it takes longer to give an answer *than* the *default-attend* strategy. On the other hand, children who still apply the wrong pre-switch rule but give correct answers due to the visual clash at the post-switch phase make decisions faster than the children who use a *strategy of control*.

Our models' results have a quite good fit to Kloo and Perner's data. However, there are also some limitations of the DCCS model. In the control group (Figure 6.5c), while the FB model's accuracy increases, the DCCS model does not improve from pre-test to post-test. This training effect of the FB model is due to the production compilation mechanism of PRIMs. The FB model initially has to retrieve each PRIM one by one in order to give an answer, therefore, it needs more time to complete the task. When the FB model is repeated, it uses the combined PRIMs and completes the task. Because we set a time threshold to the model, representing a child giving up reasoning, the accuracy of the FB model increased from pre-test to post-test even though the FB model was not trained with feedback in the control group.

On the other hand, the DCCS model has only two competing strategies to perform the task and we did not set any time threshold to the DCCS model to complete the task. The reason for this modeling choice is due to the fact that the DCCS task is not as complex as the FB task. Because the DCCS model makes an error either by selecting a wrong strategy or because of the visual clash, its accuracy does not improve from pre-test to post-test sessions in the control group after it is trained without feedback. Kloo and Perner's results showing that children's performance in the DCCS task also improved in the control group might be interpreted as getting familiar with the task itself or with the experimenter has effects on children's performance, which is not covered by our DCCS model.

To sum up, our modeling approach provides a plausible explanation for the mutual transfer between the DCCS and FB tasks and for children's errors in the DCCS and FB tasks. Based on our models, unlike Kloo and Perner (2003) and Zelazo et al. (2003), we propose that the common element in the two tasks is two competing strategies, only one of which leads to give a correct answer. Training children to use a *strategy of control* with explicit feedback explains both the training and the transfer effects.

Chapter 7: Putting the Pieces Together



Previous research on children's development of theory of mind has shown that children start to pass explicit first-order false belief tasks around the age of four²⁶ (Wellman et al., 2001). However, it takes children a couple of years after that to pass second-order false belief tasks, in which children are expected to use theory of mind recursively (Perner & Wimmer, 1985; Sullivan et al., 1994). Because it has been shown that second-order false belief reasoning is important in many different aspects of human social cognition, it is essential to understand the underlying mechanisms of children's development as well as to find efficient ways to accelerate children's development. The main goal of this dissertation was to investigate the underlying mechanisms of this developmental lag between first-order and second-order false belief reasoning and to investigate the role of feedback in accelerating children's development of second-order false belief reasoning. These questions have not been studied extensively before. In addition to this goal, we also investigated *how* working memory and cognitive control contribute to children's development of first-order false belief reasoning. To achieve these goals, we combined a computational cognitive modeling approach with empirical research. In this last chapter, I summarize our main findings. Subsequently, I discuss what these findings mean in terms of the available theories, and I provide suggestions for future work.

7.1. Summary of the results

In Chapter 1, I introduced the methodology of this dissertation (see Figure 1.3) together with the research questions. Following that methodology, in Chapter 2, I first constructed two computational cognitive models in the light of the previous research and theories on children's development of second-order theory of mind. In addition to making precise predictions that can be tested empirically, the goal of the modeling approach was to provide a procedural explanation for the research questions, "Do 5-year-olds who fail in second-order false belief tasks predominantly use zero-order theory of mind reasoning or first-order theory of mind reasoning?" and "How do 5-year-olds revise their wrong theory of mind reasoning strategy to the correct second-order theory of mind reasoning strategy over time?" To this end, we constructed computational cognitive models by using ACT-R's two possible learning mechanisms in decision making, namely *instance-based learning* and *reinforcement learning*.

The main difference between these two learning mechanisms underlies *where* and *how* strategy selection occurs. In the instance-based learning model, the reasoning strategies (i.e., zero-order, first-order, second-order) are represented

²⁶ but see Setoh, Scott, & Baillargeon (2016) for evidence that two-and-a-half-year-olds can pass explicit first-order false belief tasks when the processing demands are reduced.

as chunks in the declarative memory. Therefore, the strategy selection and revision are based on the *activation* of the strategy chunks, which are represented as declarative knowledge in the instance-based learning model. In contrast, in the reinforcement learning model, the strategy selection and revision are based on the *utilities* of the strategies, which are represented as procedural knowledge.

What does this difference between the two models mean? In the instance-based learning model, whenever a decision has to be made, the most active experience is retrieved from memory (i.e., the chunk with the highest activation) and used as the basis for the decision. The instance-based learning model revises its wrong reasoning strategy to a strategy²⁷ one level higher when it gets the feedback “Wrong” and stabilize its current strategy when it gets the feedback “Correct”. Hence, the strategy selection and revision is *explicit*. On the other hand, in the reinforcement learning model, a reward/punishment is propagated back in time through the rules that have been used to make the decision, based on feedback. This reward/punishment mechanism updates the utility of those rules and finally, the model learns to apply a correct strategy. Therefore, the reinforcement learning model selects and revises its strategy *implicitly*.

The implicit vs. explicit strategy selection led to different predictions in the two models (Figure 2.2 and Figure 2.3). Unlike the reinforcement learning model, the instance-based learning model predicted that children who fail the second-order false belief tasks but have enough experience in first-order false belief reasoning would give answers to second-order false belief questions based on first-order reasoning as opposed to zero-order reasoning. This prediction was confirmed by an empirical study that we conducted with 72 five- to six-year-old children. The results showed that 17% of the answers were correct and 83% of the answers were wrong. In line with our prediction, 65% of the wrong answers were based on a first-order theory of mind strategy, while only 29% of them were based on a zero-order strategy, and the remaining 6% was “I don’t know”.

Both models predicted that it is possible to accelerate 5-year-olds development of second-order false belief reasoning with the feedback “Wrong” without any need to explain the reasons why children’s answers are wrong. This prediction is in contrast with the previous findings on children’s development of *first-order* theory of mind, which show that it is not possible to accelerate 4-year-olds’ development of first-order theory of mind by providing feedback without explanations when they are trained on first-order false belief tasks (Clements et al., 2000; Melot & Angeard, 2003).

In addition to this prediction, the instance-based learning model predicted that providing feedback with further explanations increases the odds of

27 See the “as simple as possible as complex as necessary” argument in Chapter 2 for the reason why one level higher instead of two or more levels.

selecting the correct second-order reasoning strategy because the strategy revision is *explicit*. Therefore, if explanations are provided together with the feedback “Wrong”, children will have additional gain. On the other hand, the reinforcement learning model has nothing to say about further explanations because the strategy selection is *implicit*.

In order to test these predictions, in Chapter 3, we trained 106 5-year-old children with 12 different second-order false belief tasks in one of the following conditions: (i) Feedback with explanation; (ii) Feedback without explanations; (iii) No feedback (Figure 3.1). In the active control condition, children were trained with neutral stories that did not involve theory of mind reasoning. Confirming our instance-based learning and reinforcement learning models’ predictions, the results showed that there were significant improvements in children’s scores from pre-test to post-test in the ‘feedback without explanation’ condition (from 25% to 49%). Also, as the instance-based learning model predicted, children’s had additional gain when the feedback “Wrong” was provided with further explanations (from 31% to 68%). Moreover, surprisingly, children’s scores also improved in the ‘no feedback’ condition (from 33% to 55%). As expected, children did not show significant improvements in the ‘active control’ condition (from 29% to 35%). These improvements were not due to children’s age, verbal abilities and simple working memory span scores. Importantly, the children were able to generalize the training effect to another second-order false belief story type that they had not been trained on, and the training effect was stable at a follow-up session, which was 4 months after the pre-test.

In Chapter 2 and Chapter 3, we did not focus on the possible roles of executive functions and language in children’s development of second-order false belief reasoning. The assumption was that the children’s executive functioning and language abilities are sufficient to pass second-order false belief tasks. However, we know from previous studies that language and executive functions have an impact on children’s development of theory of mind (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Claxton, & Moses, 2014; Davis & Pratt, 1995; de Villiers & Pyers, 2002; de Villiers, 2005; de Villiers, 2007; Gordon & Olson, 1998; Keenan et al., 1998; Peterson & Siegel, 2000; Ruffman et al., 2002; Slade & Ruffman, 2005). Therefore, in Chapter 4, we conducted a cross-sectional study with 89 children in two age groups, one younger (4;6 – 6;5 years) and one older (6;7 – 8;10 years) in order to investigate the possible roles of syntactic recursion in the language domain and of working memory in the executive functions domain on children’s development of second-order theory of mind.

The reason why we focused on syntactic recursion was based on previous research showing that the syntactic component of language is related to children’s development of first-order theory of mind in terms of its hierarchical embedding

structure (de Villiers, 2005; de Villiers, 2007; Hollebrandse & Roeper, 2014) and to children's development of second-order theory of mind in terms of recursion (de Villiers et al., 2014; Hollebrandse et al., 2008). Different from those studies, for the first time in the literature, we used second-order relative clauses²⁸ in order to investigate their relation with second-order false belief reasoning. Using second-order relative clauses instead of second-order complement clauses²⁹ allowed us to specifically focus on the structural parallelism between second-order recursion in the language domain and in the thought domain by excluding the role of truth-value contrasts of complement clauses.

For the possible role of working memory in children's development of second-order false belief reasoning, we invoked the *serial processing bottleneck* hypothesis (Verbrugge, 2009), which provides a procedural account for the role of complex working memory strategies in the development of second-order false belief reasoning. While simple working memory strategies only help people to build a representation of a list of information and to report it, complex working memory strategies allow people to process information in a more efficient way with the help of combined information processing steps to perform a task. The *serial processing bottleneck* hypothesis is based on the findings that working memory acts as a bottleneck, meaning that people can only hold one chunk of information in working memory at a time (Borst et al., 2010). This hypothesis assumes that children have a time threshold to give an answer in a given task and suggests that children need complex working memory strategies in order to process embedded³⁰ beliefs in a way that chunks of information can pass through the working memory bottleneck within that time threshold.

In order to proceed in reasoning, due to the working memory bottleneck, at each step, the information in working memory needs to be sent to long-term memory to be retrieved later, if necessary. Retrieving information from long-term memory also takes time and increases the odds of forgetting and of retrieving wrong information (Anderson & Schooler, 2000). Therefore, having more inefficient rules instead of one efficient rule means that the process is more prone to errors and takes more time (Anderson et al., 2004; Taatgen & Anderson, 2002).

The evidence supporting the *serial processing bottleneck* hypothesis has been found in different cognitive domains, such as language and executive functions. Van Rij et al. (2010) focused on children's poor performance on pronoun

28 e.g., "You show me [the monkey that is pushing [a sheep that is pushing a monkey]]."

29 e.g., "John said that [Mary said that [there was a flea in her cereal]]. But in fact, she said that there was a spider in her cereal."

30 Although we explain the *serial processing bottleneck* hypothesis in terms of embedded beliefs in second-order false belief reasoning, it can be generalized to first-order and higher-orders of theory of mind reasoning. In first-order false belief reasoning, children still need to process another agent's belief which may be different from their own beliefs.

interpretation by constructing a computational model. Their model-based prediction was that once children have more time to interpret pronouns, their performance on pronoun comprehension should be increased. They validated this model-based prediction by presenting pronouns at a normal speech rate and at a slowed-down speech rate. Another empirical support to this view comes from Diamond et al.'s study (2002). They tested ninety-six 4-year-old children's executive function abilities on the day-night Stroop-like task. In this task, children were supposed to say "day" when they see a picture of the moon, and supposed to say "night" when they see a picture of the sun. They inserted several seconds of delay between the stimulus and the response by introducing a little song saying "Think about the answer; don't tell me". This manipulation improved children's performance on the task (from 56% correct to 86% correct). More recently, Ling et al. (2016) investigated whether this improvement was due to the delayed time that leads children to correct their mistakes or due to the task relevant information in the sentence "Think about the answer". To investigate these competing hypotheses, Ling et al. tested seventy-two 4-year-old children in two different conditions. In one condition the little song was saying the task-relevant information "Think about the answer; don't tell me", and in the other condition the ditty was saying the task-irrelevant information "I hope you have a nice time; I like you". The results showed that there was no difference between the task-relevant condition (83%) and the task-irrelevant condition (80%), and those conditions were significantly better than the standard version of the task (51%).

In order to test the predictions of the *serial processing bottleneck* hypothesis related to children's development of second-order false belief reasoning, we tested 89 children in two age groups, one younger (4;6 – 6;5 years) and one older (6;7 – 8;10 years) with a simple working span and a complex working span task, in addition to second-order relative clauses. The analyses showed that although second-order syntactic recursion is significantly correlated with second-order false belief reasoning, the main predictor of children's success in second-order false belief task is the complex working memory span task. Moreover, in line with the previous literature, both younger and older age groups' justification scores (i.e., answers for the question "Why?" for their judgments) were far from perfect and lower than their judgment answers for the second-order false belief questions. Most of the correct justification answers involved implicit second-order answers (e.g., "Because she doesn't know that Murat saw it") for both age groups and none of the justification answers involved explicit second-order answers (e.g., "Because she believes that Murat doesn't know that the chocolate is in the box"), except for one child in the older age group.

In addition to the highly significant effect of the complex working memory task, the simple working memory task explains significant variation in younger

children's (4 – 6) justification answers. This significant correlation of the simple working memory task disappears for older children and only the complex working memory task is able explain the variance in children's justification answers. Moreover, for the older age group, for the complex working memory task, we found that it only significantly predicts the second-order false belief justification score, not the second-order judgment score. The reason is that the judgment scores do not differ much among the older children, while their justifications still do. Thus, justifications seem to be a more sensitive variable for older children in the sense that they provide a finer distinction in their second-order reasoning abilities. While older children can give correct second-order false belief answers, their development still continues in terms of their justification abilities.

In Chapter 5 and Chapter 6, we stay with our interest in the role of executive functions but we shift our focus from children's development of second-order false belief reasoning to children's development of first-order false belief reasoning. In Chapter 5, the general research question that we wanted to investigate was "What kind of prior cognitive skills help children to pass explicit first-order false belief tasks?" More specifically, we investigated the role of simple and complex working memory strategies on children's development of first-order false belief reasoning. By constructing computational cognitive models, we aimed to simulate how children's prior skills in terms of simple and complex working memory strategies contribute to children's success in first-order false belief tasks.

In order to simulate possible transfer of skills from working memory strategies to first-order false belief reasoning, we constructed three computational cognitive models by using the cognitive architecture PRIMs, which has been built specifically to explain transfer of skills (Taatgen, 2013). Instead of choosing working memory tasks that are constructed to test children in an experimental setting, we chose two real-life tasks which children might encounter before the age of four. We named the first task the 'pencil task', in which the goal is to count the total number of yellow and green pencils in a group of blue, red, yellow and green pencils (Figure 5.2). The second task was named the 'marble task', in which the goal is to find the two bags that contain the same number of marbles of the same color (Figure 5.3). While the pencil task calls for simple working memory strategies, the marble task calls for complex working memory strategies. Finally, as a third task, we modeled a first-order false belief task. We hypothesized that performing a number of versions of the marble task would help more than the pencil task in children's transitions from failure to success in the first-order false belief task.

The simulation results showed that the first-order false belief task model learns to pass the task faster when it has prior experience of a task that needs simple or complex working memory strategies. Moreover, confirming our hypothesis, the first-order false belief task model mastered the first-order false-belief task

much faster when it was first trained in the marble task, which required complex working memory strategies, than when it was first trained in the pencil task.

In Chapter 6, we investigated another important component of executive functions, which is cognitive control in relation to children's development of first-order false belief reasoning. Previous research has shown that cognitive control and first-order false belief reasoning develop around the same age (Perner & Lang, 1999; Müller et al., 2005; Henning et al., 2011). In order to investigate whether children first develop cognitive control and then develop first-order false belief reasoning or vice versa, Kloo and Perner (2003) conducted a training study, in which 3-year-olds were trained on the Dimensional Change Card Sorting task of cognitive control in one experimental condition and in the first-order false belief tasks on the other experimental condition, both with explicit feedback with further explanations. Kloo and Perner found that there is a mutual transfer between the Dimensional Change Card Sorting task and the first-order false belief task, meaning that training children with one of the two tasks by providing feedback with further explanations significantly improved children's performance on the other task. Based on these findings, Kloo and Perner proposed that the problem that 3-year-olds encounter is related to failure in redescribing an object or situation and that training children with explicit feedback helps them to understand that an object or situation can be described differently from different perspectives. However, Kloo and Perner stated that the exact nature of the transfer effects remained to be determined.

In Chapter 6, our goal was to provide an explanation for Kloo and Perner's results by constructing computational cognitive models. To achieve this goal, we constructed our models by using the PRIMs cognitive architecture. One of the two salient explanations of transfer of skills in PRIMs theory suggests that transfer of skills between different tasks occurs when performing one of the tasks trains a particular strategy (e.g., proactive strategy), which is already present in declarative memory, and the other tasks also require to use that particular strategy or parts of that strategy in order to succeed at the task instead of using a simpler strategy (e.g., reactive strategy). Our modeling approach showed that both the Dimensional Change Card Sorting task and the first-order false belief task have a similar structure in terms of having two competing strategies, only one of which led the model to give a correct answer (Figure 5.2). Our models' results had a quite good fit to Kloo and Perner's data (Figure 5.4). Based on our models, unlike Kloo and Perner's theory, we proposed that the common element in the two tasks is two competing strategies, only one of which leads to a correct answer, namely the strategy of control. Providing children with explicit feedback trained them to use a strategy of control instead of using a simpler reactive strategy.

7.2. Discussion

In the previous section, I summarized the findings of this dissertation. In this section, I put the pieces together in the light of the theories about children's development of first-order and second-order theory of mind that I mentioned in Chapter 1, namely *conceptual change* and *complexity*. I also point out future research that will shed light on children's development of false belief reasoning.

7.2.1. Conceptual change

The conceptual change explanation for children's second-order false belief reasoning proposes that children need to realize that mental states such as beliefs can have other beliefs and not just events in the world as their content and can be used recursively (Miller 2009; 2012).

As we mentioned in Chapter 3, our findings that 5-year-old children do not need further explanations or any feedback in order to pass the second-order false belief tasks contrasts with the findings in the literature that 3-year-old children's development of first-order false belief reasoning cannot be accelerated without providing both *feedback and further explanations* when they trained on false belief tasks. Moreover, in the no feedback condition of our training study, children's second-order false belief scores also improved. This finding was surprising in terms of our models' predictions. We surmise that exposing children to second-order false belief stories and asking them second-order false belief questions, together with the justification questions "Why?" helps children to reflect about their own judgments. Thus, asking justification questions helps children to revise their wrong first-order reasoning strategy to a correct second-order reasoning strategy.

Our findings related to children's improvements in the 'feedback without explanation' and 'no feedback' conditions suggest that the conceptual change explanation alone cannot be the whole story because we did not use any second-order mental state embedding, neither in our stories nor at our training sessions. Of course, we cannot rule out the possibility that children go through a conceptual change by realizing that mental states can be used recursively between the ages three and five (see also Mahy, Moses, & Pfeifer, 2014 for a review stating that it is hard to rule out the possibility of a conceptual change at both the behavioral and neural levels).

Based on our computational modeling approach that we presented in Chapter 2, we propose that even if children go through another conceptual change after they pass the first-order false belief task, they still need experience in second-order false belief reasoning in order to revise their wrong first-order theory of mind strategy to a correct second-order strategy. Moreover, the confirmed

two predictions of our instance-based learning model indicate that children select and revise the reasoning strategies in false belief tasks explicitly.

Where do those reasoning strategies come from? A possible answer is related to learning common sense knowledge for reasoning about false beliefs. Heyes and Frith (2014) propose that explicit theory of mind is culturally inherited, and that parental stories and "causal-explanatory" statements might be some of the possible sources of this common-sense knowledge.

In the following subsection, I turn to the complexity explanation in terms of language and executive functions.

7.2.2. Complexity

The complexity explanation suggests that the higher complexity of the second-order false belief tasks adds further demands on executive functions and involves more complex language in comparison to first-order false belief tasks (Miller, 2009; 2012).

We find that understanding that mental state words can be used recursively, in spite of what the conceptual change explanation suggests, is not enough to overcome the additional demands on executive functions and to parse or use embedded clauses, such as "She *believes* that he *thinks* that the key is in the car". In order to do so, children need to process the embedded beliefs serially, which requires complex working memory strategies to pass the *serial processing bottleneck* (Verbrugge, 2009), as we argued in Chapter 4.

What might be the role of language in children's development of second-order false belief reasoning? We argue that syntactic recursion helps chunking the hierarchical embedded beliefs by linearizing them to an efficient reasoning rule that passes through the *serial processing bottleneck* of working memory (see also Hollebrandse & Roeper, 2014 for a similar argument). Our finding showing that there is a high correlation between second-order relative clauses and the complex working memory task supports this view (Chapter 4).

The *instance-based* and *reinforcement learning* models that we presented in Chapter 2 involved complex and specialized reasoning rules in the form of IF-THEN rules. However, it is unlikely that children have these complex and specialized rules in their minds to give a specific answer to false belief questions. Our computational cognitive models that we introduced in Chapter 5 showed that complex working memory strategies that involve an element of cognitive control can contribute to children's transitions from failure to success in first-order false belief tasks. Based on these findings, we propose that one of the important sources of combining those complex and specialized production rules might be children's experience in working memory strategies that they apply in their daily

lives. Future research is needed to test this explanation and the *serial processing bottleneck* hypothesis, possibly with a training study in which the children are trained with a complex working memory span task in one condition and with a simple working memory task in another condition and their performance in second-order false belief reasoning from pre-test to post-test is assessed.

In Chapter 6, our modeling approach showed us that giving explicit feedback in the Dimensional Change Card Sorting task and first-order false belief tasks train children to be more flexible in their behavior in terms of the current goals of the tasks. Because around the age four most children pass first-order false belief tasks, we can surmise that children start to have cognitive control, meaning that they learn to be flexible in their behavior depending on the current goal. Future research is needed to find out whether there is far transfer from complex working memory tasks to second-order false belief tasks. Based on our results in Chapter 3 and Chapter 4, we predict that training children with complex working memory tasks that involve not only remembering a list but also require cognitive control helps children to pass second-order false belief tasks.

However, as we have shown in Chapter 6, working memory by itself is not sufficient. Children also need to use cognitive control in order to give a correct answer. While for first-order false belief reasoning it is necessary for children to get feedback with explanations to understand that cognitive control is needed, for second-order false belief reasoning, children need to encounter many examples in which they can update their wrong strategies to the correct second-order false belief strategy.

7.3. Conclusions

In this dissertation, we combined computational cognitive modeling and empirical studies in order to investigate children's development of second-order false belief reasoning. After reviewing the previous theories, we constructed computational cognitive models based on the available theories. Constructing models gave us the opportunity to implement cognitive processes and cognitive concepts with precision instead of using the concepts without teasing apart their components. The simulation results of the models revealed novel predictions that we tested empirically. Moreover, we used computational modeling approach in order to understand the underlying mechanisms of the available data in the literature. Based on our models and empirical findings together with the previous literature, I would like to propose a timeline for the contributing factors in children's development of second-order theory of mind, which is depicted in Figure 7.1.

Development of first-order theory of mind is necessary for the development of second-order theory of mind. It seems more likely that a *conceptual change* is

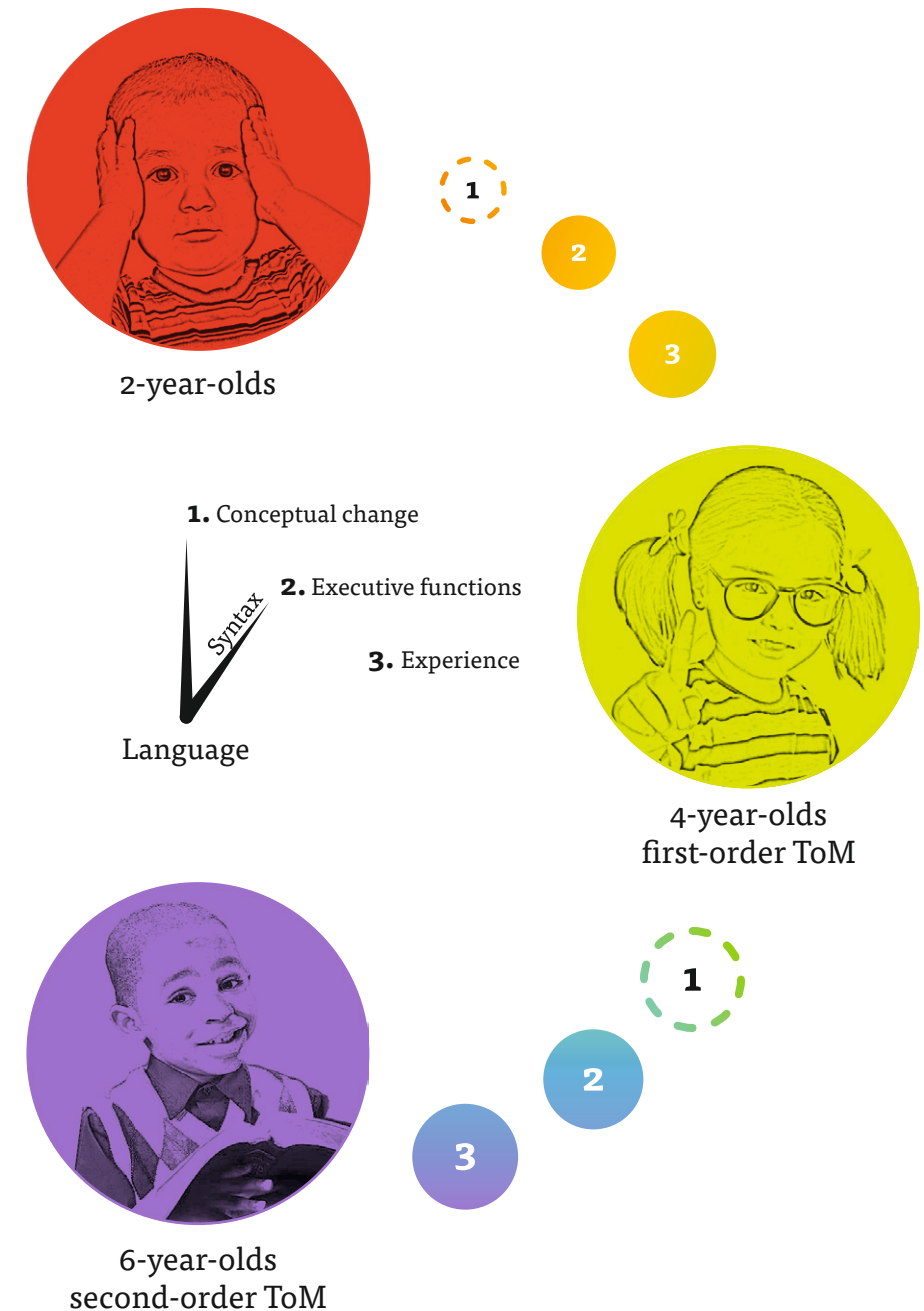


Figure 7.1. A timeline for the contributing factors in children's development of second-order theory of mind.

needed for the development of first-order theory of mind than the development of second-order theory of mind. This is because young children need first to understand conceptually others' mental states (e.g., beliefs, intentions) might be different from their own mental states. This conceptual change seems to occur with the help of pragmatics and semantics components of language.

After a possible conceptual change, young children still need to overcome the complexity of the first-order false belief tasks in terms of executive functions, including working memory and cognitive control. They need to have efficient reasoning rules to attribute false beliefs to another agent (i.e., from zero-order theory of mind reasoning to first-order theory of mind reasoning), thus they need to overcome a possible *serial processing bottleneck* of working memory. Daily life tasks that involve working memory strategies together with cognitive control contribute to having the efficient reasoning rules (Chapter 5 and Chapter 6). Moreover, the syntactic component of language helps chunking of information that can pass through a working memory bottleneck. Finally, children need experience and feedback (i.e., verbal or nonverbal) to understand that first-order theory of mind reasoning strategy is needed in a given task or situation.

After having first-order false belief understanding, children again need to overcome the complexity of the second-order false belief tasks in terms of language and working memory strategies. Syntactic recursion in the language domain is a right representational tool to help them to have efficient reasoning rules to process hierarchical embedded beliefs in a linear way, in order to pass the *serial processing bottleneck* of working memory (Chapter 4).

After being able to deal with the complex working memory strategies, children still need experience in second-order false belief reasoning in order to revise their wrong first-order reasoning strategy to the correct second-order reasoning strategy (Chapter 2 and Chapter 3). This experience can be gained through reading narrative fiction and playing games that require second-order theory of mind as well as through social communication with parents, friends and siblings.

Samenvatting

Wat zou ik moeten schrijven om uw aandacht te trekken, en u er van te overtuigen om dit hele proefschrift te lezen, of in ieder geval tot het einde van deze samenvatting? Om dit doel te bereiken zou ik allereerst in uw schoenen moeten stappen door na te denken over uw achtergrond. Daarna zou ik een manier moeten vinden om u genoeg informatie te geven om de inhoud van dit proefschrift te begrijpen, zonder u daarbij van verveling in slaap te laten vallen. In andere woorden, ik zou mijn *theory of mind* (Premack & Woodruff, 1978) moeten gebruiken, en me realiseren dat u misschien andere kennis, ideeën en doelen heeft dan ik.

Misschien *denkt* u dat ik *geloof* dat ik een manier heb gevonden om uw aandacht te trekken, maar *denkt* u ook dat ik me daarin vergis. Wanneer ik nadenk over het feit dat “u *denkt* dat ik het mis heb”, maak ik gebruik van eerste-orde *theory of mind*, door een mentale toestand aan u toe te schrijven. Daarnaast maak ik gebruik van tweede-orde *theory of mind* wanneer ik nadenk over het feit dat “u *denkt* dat ik *geloof* dat ik een manier heb gevonden”, door aan u een gedachte toe te schrijven die gaat over een mentale toestand van mij. Het toeschrijven van dit soort mentale toestanden van de tweede orde is belangrijk in vele sociale situaties, waaronder het begrijpen van idiomen (Caillies & Le Sourn-Bissaoui, 2013), het volhouden van een leugen (Hsu & Cheung, 2013), en het begrijpen van ironie (Filippova & Astington, 2008). Dit speelt bijvoorbeeld een rol wanneer John zegt “Je bent een geweldige onderzoeker”, terwijl John niet de bedoeling heeft dat Stefan denkt dat hij een goede onderzoeker is.

Deze dissertatie is onderdeel van een project genaamd “Cognitive systems in interaction: Logical and computational models of higher-order social cognition”, dat is toegekend aan mijn eerste begeleider Rineke Verbrugge. Het overkoepelende doel van dit project is om een beter begrip te verkrijgen voor hogere-orde *theory of mind* ten behoeve van cognitiewetenschappers, logici en informatici. In de nabije toekomst zullen mensen in hun dagelijkse leven samenwerken met kunstmatige agents. Door de onderliggende mechanismen en beperkingen van het hogere-orde redeneren van mensen te onderzoeken, kunnen we systemen bouwen voor effectievere communicatie, samenwerking en onderhandelingen tussen mens en machine. Ik heb me gericht op de manier waarop tweede-orde *theory of mind* zich ontwikkelt in kinderen, in termen van het maken van beslissingen, transfer of skills, cognitieve controle, werkgeheugen en taal.

Naast de bijdrage die het levert aan ontwikkelingspsychologie en cognitiewetenschappen in het algemeen, kan onderzoek naar het ontwikkelpatroon van *theory of mind* in kinderen bijdragen aan het nieuwe veld van kind-robot-interacties. Kind-robot-interacties zijn bedoeld om de gezondheidszorg en educatie van kinderen te verbeteren met behulp van interactieve robots. Aangezien

kinderen robots niet zien als programma's, maar eigenschappen van levende wezens toeschrijven aan robots (Belpaeme et al., 2013), is het belangrijk dat de robots die met kinderen interacteren "weten" wat de beperkingen zijn van de theory of mind van een kind gegeven de leeftijd van dat kind, maar ook wat de onderliggende mechanismen zijn voor deze beperkingen. Daarnaast is het voor de sociale en cognitieve vaardigheden van de robot belangrijk dat de robot effectieve manieren "weet" om kinderen te stimuleren in hun theory of mind. Deze dissertatie heeft als doel om bij te dragen aan alle bovengenoemde onderzoeksgebieden door nieuwe inzichten te presenteren over de ontwikkeling van tweede-orde theory of mind in kinderen.

Eerder onderzoek naar de ontwikkeling van theory of mind in kinderen heeft aangetoond dat kinderen rond hun vierde¹ jaar beginnen te slagen in de eerste-orde false-belief-taak (Wellman et al., 2001). Het kost kinderen echter nog enkele jaren voordat ze ook slagen in de tweede-orde false-belief-taak, waarin kinderen hun theory of mind recursief moeten toepassen (Perner & Wimmer, 1985; Sullivan et al., 1994). Omdat tweede-orde theory of mind belangrijk is in verschillende aspecten van menselijke sociale cognitie is het essentieel om de onderliggende mechanismen van de ontwikkeling van theory of mind in kinderen te begrijpen en om efficiënte manieren te vinden om deze ontwikkeling te versnellen. Het hoofddoel van deze dissertatie is om de onderliggende mechanismen te onderzoeken die de tijd verklaren tussen de ontwikkeling van eerste-orde theory of mind en tweede-orde theory of mind en om te bepalen welke rol feedback speelt in het versnellen van de ontwikkeling van tweede-orde theory of mind in kinderen. Er is nog weinig aandacht geweest voor deze vragen. Daarnaast onderzoeken we ook hoe werkgeheugen en cognitieve controle bijdrage aan de ontwikkeling van eerste-orde theory of mind in kinderen. Hiervoor maken we gebruik van een combinatie van computationele cognitieve modellen en empirisch onderzoek.

In Hoofdstuk 1 introduceer ik de methodologie en de onderzoeksvragen in dit proefschrift (zie Figuur 1.3). In navolging van deze methodologie presenteert Hoofdstuk 2 twee computationele cognitieve modellen die ik heb gebouwd aan de hand van eerder onderzoek en theorieën over de ontwikkeling van tweede-orde theory of mind in kinderen. Het doel van deze modelleeraanpak was, naast het doen van exacte voorspellingen die empirisch getest kunnen worden, om een procedurele verklaring te geven voor de onderzoeksvraag "Gebruiken kinderen van vijf jaar oud die niet slagen in de tweede-orde false-belief-taak voornamelijk nulde-orde of eerste-orde theory of mind?" Om deze vraag te beantwoorden zijn computationele cognitieve modellen gebouwd met behulp van de twee mogelijke leermechanismen in ACT-R, te weten *instance-based* leren en *reinforcement* leren.

¹ Setoh, Scott, & Baillargeon (2016) presenteren daarentegen aanwijzingen dat kinderen van twee-en-een-half jaar oud expliciete eerste-orde false-belief-taken kunnen doen wanneer de taken laagdrempeliger zijn.

Het voornaamste verschil tussen deze twee leermechanismen geeft aan waar en hoe strategieselectie plaatsvindt. In het model met instance-based leren worden redeneerstrategieën (d.w.z. nulde-orde, eerste-orde, tweede-orde theory of mind) gerepresenteerd als chunks in het declaratieve geheugen. Strategieselectie en -revisie zijn daarom gebaseerd op de *activatie* van deze strategie-chunks. In het model met reinforcement leren, daarentegen, zijn strategieselectie en -revisie gebaseerd op *utilities* van de strategieën, die zijn gerepresenteerd als procedurele kennis.

Wat betekent dit verschil tussen de twee modellen? In het model van instance-based leren geldt dat wanneer er een beslissing gemaakt moet worden, de meest actieve ervaring in het geheugen (d.w.z. de chunk met de hoogste activatie) de basis is voor de beslissing. Wanneer deze beslissing wordt gevolgd door de feedback "Fout", past het instance-based leermodel de strategie naar de strategie² die één orde theory of mind hoger is. Bij de feedback "Correct" stabiliseert het gebruik van de gekozen strategie zich. Strategieselectie en -revisie zijn daarom *expliciet*. Daarentegen wordt in het reinforcement leermodel een beloning of straf teruggevoerd naar de regels die ten grondslag lagen aan de beslissing die leidde tot deze beloning of straf. Dit mechanisme van beloningen en straffen past de utilities van deze regels aan, waardoor het model uiteindelijk leert de juiste strategie te hanteren. In het reinforcement leermodel zijn strategieselectie en -revisie daarom *impliciet*. Het verschil tussen impliciete en expliciete strategieselectie leidt tot voorspellingen die verschillen voor de twee modellen (Figuur 2.2 en Figuur 2.3). In tegenstelling tot het reinforcement leermodel, voorspelt het instance-based leermodel dat kinderen die een fout antwoord geven in de tweede-orde false-belief-taak een antwoord geven op basis van eerste-orde theory of mind redeneren, als ze hierin voldoende ervaring hebben opgedaan. Deze voorspelling is door ons bevestigd in empirisch onderzoek met 72 kinderen in de leeftijd van vijf tot zes jaar oud. De resultaten laten zien dat 17% van de antwoorden correct waren tegen 83% foute antwoorden. In lijn met onze voorspelling werd in 65% van de foute antwoorden een antwoord gegeven dat overeenkomt met de eerste-orde theory of mind strategie, terwijl 29% overeenkomst met de nulde-orde theory of mind strategie. De overige 6% van de antwoorden was "Ik weet het niet".

Beide modellen voorspellen dat het mogelijk is om de ontwikkeling van tweede-orde theory of mind te versnellen in kinderen van vijf jaar oud door de feedback "Fout" te geven zonder de reden aan te geven waarom het antwoord van het kind incorrect is. Deze voorspelling is in tegenstelling met eerdere resultaten in de ontwikkeling van eerste-orde theory of mind in kinderen, waaruit blijkt dat het

² Zie ook het "zo simpel mogelijk en zo complex als noodzakelijk" argument in Hoofdstuk 2 voor de reden waarom de strategie één orde theory of mind wordt verhoogd in plaats van twee of meer.

niet mogelijk is om de ontwikkeling van eerste-orde theory of mind in kinderen te versnellen door feedback zonder uitleg te geven in de eerste-orde false-belief-taak (Clements et al., 2000; Melot & Angeard, 2003).

Daarnaast voorspelt het instance-based leermodel dat het geven van feedback met uitleg de kansen vergroot om de juiste tweede-orde theory of mind strategie te selecteren omdat strategierevisie expliciet is. Dat wil zeggen, kinderen hebben er baat bij om uitleg te krijgen voor het feit dat ze de feedback “Fout” hebben gekregen. Het reinforcement leermodel doet daarentegen geen uitspraken over uitleg van feedback omdat strategieselectie impliciet is.

Om deze voorspellingen in Hoofdstuk 3 te testen hebben we 106 kinderen van 5 jaar oud getraind met 12 verschillende tweede-orde false-belief-taken in de volgende condities: (i) Feedback met uitleg; (ii) Feedback zonder uitleg; (iii) Geen feedback (Figuur 3.1). In de actieve controleconditie werden kinderen getraind met neutrale verhalen waarvoor geen theory of mind nodig was. De resultaten laten zien dat kinderen significant vaker correcte antwoorden geven in de post-test dan de pre-test in de ‘feedback zonder uitleg’ conditie (van 25% naar 49%), en bevestigen daarmee de voorspellingen van onze instance-based en reinforcement leermodellen. Zoals voorspeld door het instance-based leermodel hadden kinderen daarnaast baat bij extra uitleg wanneer ze de feedback “Fout” ontvingen (van 31% naar 68%). Verrassenderwijs gaven kinderen ook vaker het juiste antwoord in de ‘geen feedback’ conditie (van 33% naar 55%). Zoals verwacht gaven kinderen niet significant vaker het juiste antwoord in de actieve controleconditie (van 29% naar 35%). Deze vooruitgang na training met tweede-orde false-belief-taken kan niet worden verklaard aan de hand van de leeftijd, verbale vermogen of de scores op simpele werkgeheugentaken. Daarnaast konden de kinderen het effect van hun training generaliseren naar een ander type tweede-orde false-belief-taak waar ze niet op waren getraind. Dit trainingseffect was stabiel bij een vervolgsessie, 4 maanden na de pre-test.

In Hoofdstuk 2 en Hoofdstuk 3 hebben we niet gekeken naar de mogelijke rol van executive functies en taal in de ontwikkeling van tweede-orde theory of mind in kinderen. De aanname was dat kinderen voldoende vaardigheden op het gebied van executive functies en taal hebben om te slagen in tweede-orde false-belief-taken. Eerder onderzoek laat echter zien dat taal en executive functies effect kunnen hebben op de ontwikkeling van theory of mind in kinderen (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Claxton, & Moses, 2014; Davis & Pratt, 1995; de Villiers & Pyers, 2002; de Villiers, 2005; de Villiers, 2007; Gordon & Olson, 1998; Keenan et al., 1998; Peterson & Siegel, 2000; Ruffman et al., 2002; Slade & Ruffman, 2005). In Hoofdstuk 4 presenteren we daarom een transversaal onderzoek met 89 kinderen, verdeeld over twee leeftijdsgroepen, een jongere groep (4;6 – 6;5 jaar oud) en een oudere groep (6;7 – 8;10 jaar oud),

om de mogelijke rol van syntactische recursie in het taaldomein en van werkgeheugen in het domein van executive functie op de ontwikkeling van tweede-orde theory of mind in kinderen te onderzoeken.

De reden voor het gebruik van syntactische recursie is gebaseerd op eerder onderzoek dat laat zien dat de syntactische component van taal gerelateerd is aan de ontwikkeling van eerste-orde theory of mind in kinderen in termen van hiërarchische inbeddingsstructuur (de Villiers, 2005; de Villiers, 2007; Hollebrandse & Roeper, 2014) en aan de ontwikkeling van tweede-orde theory of mind in termen van recursie (de Villiers et al., 2014; Hollebrandse et al., 2008). In tegenstelling tot deze eerdere onderzoeken hebben wij gebruik gemaakt van tweede-orde bijvoeglijke bijzinnen³ om hun relatie met tweede-orde false-belief-taken te onderzoeken. Omdat we tweede-orde bijvoeglijke bijzinnen gebruiken in plaats van tweede-orde zelfstandige bijzinnen⁴ kunnen we ons specifiek richten op structurele parallellen tussen tweede-orde recursie in het taaldomein en het mentale domein door de rol van waarheid van de bijzinnen uit te sluiten.

Om de rol van werkgeheugen in de ontwikkeling van tweede-orde theory of mind te onderzoeken hebben we gebruik gemaakt van de *serial processing bottleneck* hypothese (Verbrugge, 2009), die een procedurele beschrijving geeft van de rol van complexe werkgeheugenstrategieën in de ontwikkeling van tweede-orde theory of mind. Waar simpele werkgeheugenstrategieën mensen alleen helpen met het opbouwen van een representatie van een lijst informatie en hierover te rapporteren, kunnen mensen met behulp van complexe werkgeheugenstrategieën informatie op een efficiëntere manier verwerken door verschillende stappen in de informatieverwerking te combineren. De *serial processing bottleneck* hypothese is gebaseerd op het resultaat dat werkgeheugen een beperkende factor is in het verwerken van informatie, doordat mensen slechts één chunk informatie tegelijk kunnen verwerken (Borst et al., 2010). Deze hypothese neemt aan dat kinderen een tijdslimiet hebben voor het uitvoeren van een bepaalde taak, en suggereert dat kinderen complexe werkgeheugenstrategieën nodig hebben om ervoor te zorgen dat hogere-orde⁵ mentale toestanden kunnen worden verwerkt binnen deze tijdslimiet.

Vanwege de beperking in het werkgeheugen moet tijdens elke stap in het rede-neerproces informatie van het werkgeheugen worden opgeslagen in het langetermijngeheugen om later, indien nodig, weer opgehaald te worden. Informatie ophalen uit het langetermijngeheugen kost tijd en verhoogt de kans dat informatie

3 bijv., “Je wijst naar de aap [die duwt tegen het schaap [die duwt tegen een aap]].”

4 bijv., “John zei [dat Mary zei [dat er een vlieg in haar soep zat]]. Maar er zat een spin in haar soep.”

5 Hoewel we de *serial processing bottleneck* hypothese in termen van hogere-orde mentale toestanden beschrijven, kan dit worden gegeneraliseerd naar eerste-orde en hogere-orde theory of mind. In de eerste-orde false-belief-tak hebben kinderen als taak om te onthouden wat een andere agent denkt dat de situatie is, ook als het kind er zelf anders over denkt.

wordt vergeten of dat verkeerde informatie wordt opgehaald (Anderson & Schooler, 2000). Hieruit volgt dat een proces dat bestaat uit meerdere inefficiënte regels vaker tot fouten leidt en meer tijd kost dan een proces dat slechts één efficiënte regel heeft (Anderson et al., 2004; Taatgen & Anderson, 2002).

De *serial processing bottleneck* hypothese wordt bevestigd door resultaten uit verschillende cognitieve domeinen, zoals taal en executive functies. Van Rij et al. (2010) richtten zich op moeilijkheden die kinderen ondervinden bij het produceren van voornaamwoorden door een computationeel model te bouwen. De modelvoorspelling was dat als kinderen meer tijd zouden hebben om voornaamwoorden te interpreteren, hun prestaties op het begrip van voornaamwoorden zouden moeten verbeteren. Deze voorspelling hebben ze gevalideerd door voornaamwoorden te presenteren op normale spreeknelheid en op een vertraagde spreeknelheid. Empirische ondersteuning komt ook vanuit het onderzoek van Diamond et al. (2002). Ze testten de vaardigheden op gebied van executive functies van 96 kinderen van 4 jaar oud door middel van de dag-nacht taak. In deze taak, die vergelijkbaar is aan de Stroop taak, moesten kinderen “dag” zeggen wanneer ze een afbeelding van de maan zagen, en “nacht” zeggen als ze een afbeelding van de zon zagen. Ze dwongen een aantal seconden pauze af tussen stimulus en respons door een kort liedje te spelen dat zei “Denk na over het antwoord, zeg het nog niet”. Deze manipulatie zorgde voor een verbetering in de prestaties van kinderen (van 56% naar 86% juiste responses). In recenter werk onderzochten Ling et al. (2016) of deze verbetering een gevolg was van de extra tijd, waardoor kinderen hun fouten konden herstellen, of door de zin “Denk na over het antwoord”. Ling et al. testten 72 kinderen van 4 jaar oud in twee condities. In de eerste conditie gaf het liedje informatie over de taak door de zin “Denk na over het antwoord, zeg het nog niet”, terwijl het liedje in de andere conditie de tekst “Ik hoop dat je je vermaakt, ik vind je aardig” was. De resultaten lieten geen verschil zien tussen de tekst die relevant was voor de taak (83%) en tekst die niet relevant was (50%). Beide versies waren significant beter dan de standaardversie van de taak (51%).

Om de voorspellingen van de *serial processing bottleneck* hypothese op het gebied van de ontwikkeling van tweede-orde theory of mind in kinderen te testen, hebben we 89 kinderen in twee leeftijdsgroepen, een jongere (4;6 – 6;5 jaar) en een oudere (6;7 – 8;10 jaar), getest met een simpele werkgeheugentaak en een complexe werkgeheugentaak naast een test op hun begrip van tweede-orde bijvoeglijke bijzinnen. De analyses laten zien dat, hoewel tweede-orde syntactische recursie significant correleert met redeneren aan de hand van tweede-orde theory of mind, de prestatie op de complexe werkgeheugentaak een betere voorspelling doet over het al dan niet slagen van kinderen in de tweede-orde false-belieftaak. Daarnaast waren, in lijn met bevindingen in de literatuur, de

verantwoordingsscores (d.w.z. antwoorden op de vraag “Waarom?”) van zowel de jongere als de oudere leeftijdsgroep verre van perfect en lager dan de onuitgelegde antwoordscores op de tweede-orde false-belieftaak. Voor beide leeftijdsgroepen waren correcte verantwoordingen voornamelijk van de vorm van impliciete tweede-orde antwoorden (bijv. “Omdat ze niet weet dat Murat het zag”), en waren er geen expliciete tweede-orde verantwoordingen (bijv. “Omdat ze denkt dat Murat niet weet dat de chocola in de doos zit”), met uitzondering van één kind in de oudere leeftijdsgroep.

Naast het significante effect van de complexe werkgeheugentaak, verklaart de prestatie op de simpele werkgeheugentaak een significante deel van de variatie in de verantwoordingsscores van jongere kinderen (4 – 6 jaar oud). Deze significante correlatie met de prestaties op de simpele werkgeheugentaak verdwijnt voor de oudere kinderen, voor wie alleen de prestaties op de complexe werkgeheugentaak variatie in de verantwoordingsscores verklaart. Bovendien laten we zien dat de prestaties op de complexe werkgeheugentaak voor de oudere leeftijdsgroep alleen de verantwoordingsscores significant verklaart, maar niet de scores op de tweede-orde false-belieftaak. De reden hiervoor is dat er weinig variatie is in de onuitgelegde antwoordscores op de false-belieftaak in deze leeftijdsgroep, terwijl er nog wel variatie is in der verantwoordingsscores. Verantwoording van antwoorden lijkt daarom een belangrijke variabele te zijn, omdat het een betere scheiding geeft van de tweede-orde redeneervaardigheden van kinderen. Hoewel oudere kinderen correcte antwoorden kunnen geven in de tweede-orde false-belieftaak, ontwikkelen deze kinderen zich nog steeds op het gebied van de verantwoording van hun antwoorden.

Hoofdstuk 5 en Hoofdstuk 6 gaan door op de rol van executive functies, maar verschuiven de aandacht van de ontwikkeling die kinderen doormaken op het gebied van de tweede-orde false-belieftaak naar de ontwikkeling die ze doormaken op het gebied van de eerste-orde false-belieftaak. In Hoofdstuk 5 staat de onderzoeksvraag “Welk soort cognitieve vaardigheden helpt kinderen in het slagen in de expliciete eerste-orde false-belieftaak?” centraal. Hiervoor onderzochten we de rol van simpele en complexe werkgeheugenstrategieën op de ontwikkeling van kinderen op gebied van de eerste-orde false-belieftaak. Ons doel was om te simuleren hoe de cognitieve vaardigheden van kinderen in termen van simpele en complexe werkgeheugenstrategieën bijdragen aan het succes in de eerste-orde false-belieftaak door het bouwen van een computationeel cognitief model.

Om de mogelijke bijdrage van werkgeheugenstrategieën aan eerste-orde theory of mind redeneren te simuleren hebben we drie computationele cognitieve modellen gebouwd in de cognitieve architectuur PRIMs, die speciaal gebouwd is om dit soort bijdrages te verklaren (Taatgen, 2013). In plaats van werkgeheugentaken te modelleren die kinderen testen in een laboratoriumomgeving, hebben

we ervoor gekozen om twee taken te modelleren die kinderen in het dagelijkse leven zouden kunnen tegenkomen voordat ze vier jaar oud zijn. De eerste taak hebben we de 'potloodtaak' genoemd, waarin het doel is om het totaal aantal gele en groene potloden te tellen in een groep van blauwe, rode, gele en groene potloden (Figuur 5.2). Het doel van de tweede taak, genaamd de 'knikkertaak', is om twee zakken te vinden waarin hetzelfde aantal knikkers van dezelfde kleur zit (Figuur 5.3). Waar de potloodtaak alleen om simpele werkgeheugenstrategieën vraagt, zijn voor de knikkertaak ook complexe werkgeheugenstrategieën nodig. Als derde taak hebben we een eerste-orde false-belieftaak gemodelleerd. Onze hypothese was dat ervaring met de knikkertaak meer zou bijdragen aan succes in de eerste-orde false-belieftaak dan ervaring met de potloodtaak.

De simulatieresultaten laten zien dat ervaring met taken die vragen om zowel simpele als complexe werkgeheugenstrategieën bijdragen aan succes in de eerste-orde false-belieftaak. Daarnaast is onze hypothese bevestigd door het feit dat het model de eerste-orde false-belieftaak veel sneller leerde wanneer het werd getraind met de knikkertaak, waarvoor complexe werkgeheugenstrategieën nodig zijn, dan wanneer het werd getraind met de potloodtaak.

In Hoofdstuk 6 onderzoeken we een andere belangrijke component van executive functies, namelijk cognitieve controle, in relatie tot de ontwikkeling van eerste-orde theory of mind redeneren in kinderen. Eerder onderzoek heeft aangetoond dat cognitieve controle en eerste-orde theory of mind redeneren zich rond dezelfde leeftijd ontwikkelen in kinderen (Perner & Lang, 1999; Müller et al., 2005; Henning et al., 2011). Om te bepalen of kinderen eerst cognitieve controle ontwikkelen en daarna eerste-orde theory of mind redeneren of vice versa, voerden Kloo en Perner (2003) een trainingsstudie uit, waarin kinderen van drie jaar oud werden getraind op de Dimensional Card Sorting taak, waarvoor cognitieve controle nodig is, of de eerste-orde false-belieftaak. In beide condities werd expliciete feedback gegeven met uitleg over het juiste antwoord. De resultaten van Kloo en Perner laten zien dat er wederzijds een bijdrage is van de score op de Dimensional Card Sorting taak enerzijds en de eerste-orde false-belieftaak anderzijds, wat betekent dat kinderen die worden getraind op één van beide taken significant beter scoren op de andere taak. Op basis van deze bevindingen concludeerden Kloo en Perner dat kinderen van 3 jaar oud mogelijk problemen ondervinden om de representatie van een object of situatie te veranderen, en dat trainen met expliciete feedback kinderen helpt te begrijpen dat een object of situatie verschillend kan worden beschreven, afhankelijk van het perspectief. Kloo en Perner merkten echter ook op dat de exacte bijdrage van de ene taak op de andere taak nog onbekend is.

Ons doel in Hoofdstuk 6 is om een verklaring te geven voor de resultaten van Kloo en Perner door een computationeel cognitief model te bouwen in de

cognitieve architectuur PRIMs. PRIMs theorie claimt dat trainen op een taak kan bijdragen aan het succes op een andere taak wanneer in beide taken succes afhangt van het gebruik van dezelfde onderliggende strategie (bijv. een proactieve strategie) in het declaratief geheugen. Trainen op de ene taak kan dan voorkomen dat een simpelere strategie (bijv. een reactieve strategie) wordt geselecteerd om de andere taak uit te voeren. Ons model laat zien dat de Dimensional Card Sorting taak en de eerste-orde false-belieftaak een gezamenlijke onderliggende structuur hebben, waarin twee strategieën met elkaar in competitie zijn, waarvan slechts één leidt tot het juiste antwoord (Figuur 5.2). De resultaten van ons modellen passen goed bij de data van Kloo en Perner (Figuur 5.4). Op basis van deze modellen concluderen we dat, in tegenstelling tot wat Kloo en Perner beweren, de competitie tussen strategieën, waarvan slechts één leidt tot het juiste antwoord, het gemeenschappelijke element is in de twee taken. Door kinderen expliciet feedback te geven worden ze getraind in het gebruik van de controlestrategie die leidt tot het juiste antwoord, in plaats van te vertrouwen op een simpelere reactieve strategie.

Discussie

In de vorige sectie heb ik de resultaten van dit proefschrift samengevat. In deze sectie voeg ik de delen samen en bekijk het geheel in het licht van de theorieën uit Hoofdstuk 1 over de ontwikkeling van eerste-orde en tweede-orde theory of mind in kinderen, namelijk *nieuw inzicht* en *complexiteit*. Ik noem ook mogelijkheden voor vervolgonderzoek die meer licht zouden kunnen werpen over de ontwikkeling van de theory of mind van kinderen.

Nieuw inzicht

De *nieuwe inzicht* verklaring claimt dat kinderen alleen kunnen slagen voor de tweede-orde false-belieftaak wanneer ze zich realiseren dat mentale toestanden zoals overtuigingen bestaan, dat iemand overtuigd kan zijn van iets dat niet waar is, en dat ze deze concepten recursief kunnen toepassen (Miller 2009; 2012).

Zoals we hebben opgemerkt in Hoofdstuk 3, zijn onze resultaten dat kinderen die feedback zonder extra uitleg krijgen sneller leren te slagen in de tweede-orde false-belieftaak in tegenstelling met de literatuur die aantoonde dat training in de eerste-orde false-belieftaak niet helpt kinderen van 3 jaar oud sneller te laten slagen voor de eerste-orde false-belieftaak, tenzij ze *feedback en extra uitleg* krijgen. Daarnaast verbeterden de scores op de tweede-orde false-belieftaak ook in de

conditie waarin ze geen feedback kregen. Dit resultaat was opmerkelijk gegeven onze modelvoorspellingen. We veronderstellen dat door blootstelling aan tweede-orde false-beliefverhalen en door tweede-orde theory of mind vragen te stellen met de verantwoordingsvraag “Waarom?”, kinderen worden gestimuleerd om na te denken over hun eigen antwoorden. Dat wil zeggen, door de verantwoordingsvraag te stellen worden kinderen gestimuleerd om hun incorrecte eerste-orde redeneerstrategie te corrigeren naar een tweede-orde redeneerstrategie.

De verbetering in prestatie die we zien in de conditie ‘feedback zonder verklaring’ en de conditie ‘geen feedback’ suggereert dat een nieuw inzicht niet de enige verklaring kan zijn, omdat we geen tweede-orde inbedding hebben gebruikt voor mentale toestanden in de verhalen of in de trainingssessies. Natuurlijk kunnen we niet uitsluiten dat kinderen nieuw inzicht verkrijgen over het recursief redeneren over mentale toestanden als ze drie en vijf jaar oud zijn (zie ook Mahy, Moses, & Pfeifer, 2014 voor een review die claimt dat het moeilijk is om de mogelijkheid van nieuw inzicht uit te sluiten, zowel op gedragsniveau als op neurale niveau).

Gezien de resultaten van onze computationele modelleeraanpak van Hoofdstuk 2, stellen we dat zelfs als kinderen een nieuw inzicht verkrijgen nadat ze slagen in de eerste-orde false-belieftaak, ze ook dan nog ervaring moeten krijgen in tweede-orde theory of mind redeneren om hun eerste-orde redeneerstrategie te verbeteren tot een tweede-orde redeneerstrategie. Daarnaast duiden de twee bevestigde voorspellingen van ons instance-based leermodel er op dat kinderen hun redeneerstrategieën expliciet selecteren en aanpassen.

Waar komen deze redeneerstrategieën vandaan? Een mogelijk antwoord is dat het ontwikkelen van theory of mind gerelateerd is aan het leren van algemene regels over gedrag. Heyes en Frith (2014) stellen dat expliciete theory of mind wordt doorgegeven in de cultuur, en dat kinderverhalen en “causaal-verklarende” uitspraken mogelijke bronnen zijn van algemene kennis over gedrag van mensen.

In de volgende sectie bespreek ik de complexiteitsverklaring in termen van taal en executive functies.

Complexiteit

De complexiteitsverklaring claimt dat voor de complexere tweede-orde false-belieftaak een grotere vaardigheid op het gebied van executive functies en complexe taal nodig is vergeleken met de eerste-orde false-belieftaak (Miller, 2009; 2012).

Onze resultaten laten zien dat het begrip dat woorden die mentale toestanden beschrijven recursief kunnen worden gebuikt, in tegenstelling tot wat de nieuwe-inzichtverklaring beweert, niet genoeg is om de extra vraag naar executive

functies te beperken om bijzinnen zoals in “Zij denkt dat hij denkt dat de sleutel in de auto ligt” te verwerken. Om dit soort zinnen te verwerken moeten kinderen de ingebedde mentale toestanden serieel verwerken, waarvoor kinderen vanwege de *serial processing bottleneck* complexe werkgeheugenstrategieën nodig hebben (Verbrugge, 2009), zoals besproken in Hoofdstuk 4.

Wat zou de rol van taal kunnen zijn in de ontwikkeling van tweede-orde theory of mind redeneren in kinderen? Volgens ons helpt syntactische recursie met het verdelen van hiërarchisch ingebedde mentale toestanden in chunks door ze serieel te maken aan de hand van een efficiënte redeneerregel die makkelijk door de *serial processing bottleneck* gaat (zie ook Hollebrandse & Roeper, 2014 voor een vergelijkbaar argument). Dit wordt ondersteund door onze resultaten, die laten zien dat er een hoge correlatie is tussen het begrip van tweede-orde bijvoeglijke bijzinnen en de score op de complexe werkgeheugentaak (Hoofdstuk 4).

De *instance-based* en de *reinforcement* leermodellen uit Hoofdstuk 2 bevatten complexe en gespecialiseerde redeneerregels in de vorm van ALS-DAN regels. Het is echter onwaarschijnlijk dat kinderen deze complexe en gespecialiseerde regels expliciet gebruiken om een specifiek antwoord te geven in de false-belieftaak. De computationele cognitieve modellen die we hebben geïntroduceerd in Hoofdstuk 5 laten zien dat complexe werkgeheugenstrategieën die een element van cognitieve controle bevatten bij kunnen dragen aan het slagen van kinderen in de eerste-orde false-belieftaak. Op basis van deze resultaten stellen we dat de ervaring die kinderen in het dagelijks leven opdoen met werkgeheugenstrategieën één van de belangrijke bronnen zou kunnen zijn van deze complexe en gespecialiseerde productieregels. Om deze verklaring en de *serial processing bottleneck* hypothese te testen is meer onderzoek nodig, mogelijk een trainingsstudie waarin kinderen getraind worden met een complexe werkgeheugentaak in de ene conditie en met een simpele werkgeheugentaak in de andere conditie, waarbij hun vooruitgang op het gebied van tweede-orde theory of mind redeneren van pre-test tot post-test wordt geëvalueerd.

In Hoofdstuk 6 liet onze modelleeraanpak zien dat het geven van expliciete feedback in de Dimensional Card Sorting taak en de eerste-orde false-belieftaak kinderen traint om flexibeler te zijn in hun gedrag in termen van de huidige doelen van een taak. Omdat de meeste kinderen slagen in de eerste-orde false-belieftaak rond het vierde levensjaar, vermoeden we dat kinderen dan cognitieve controle beginnen uit te oefenen, wat betekent dat ze leren om flexibel hun gedrag aan te passen aan hun huidige doel. Om te bepalen of er ook een bijdrage is van complexe werkgeheugenstrategieën aan het slagen in de tweede-orde false-belieftaak (*far transfer*) is meer onderzoek nodig. Op basis van onze resultaten in Hoofdstuk 3 en Hoofdstuk 4 voorspellen we dat de prestaties van kinderen op de tweede-orde false-belieftaak verbeteren door training met complexe

werkgeheugentaken, die niet alleen bestaan uit het herinneren van een lijst maar waarvoor ook cognitieve controle nodig is.

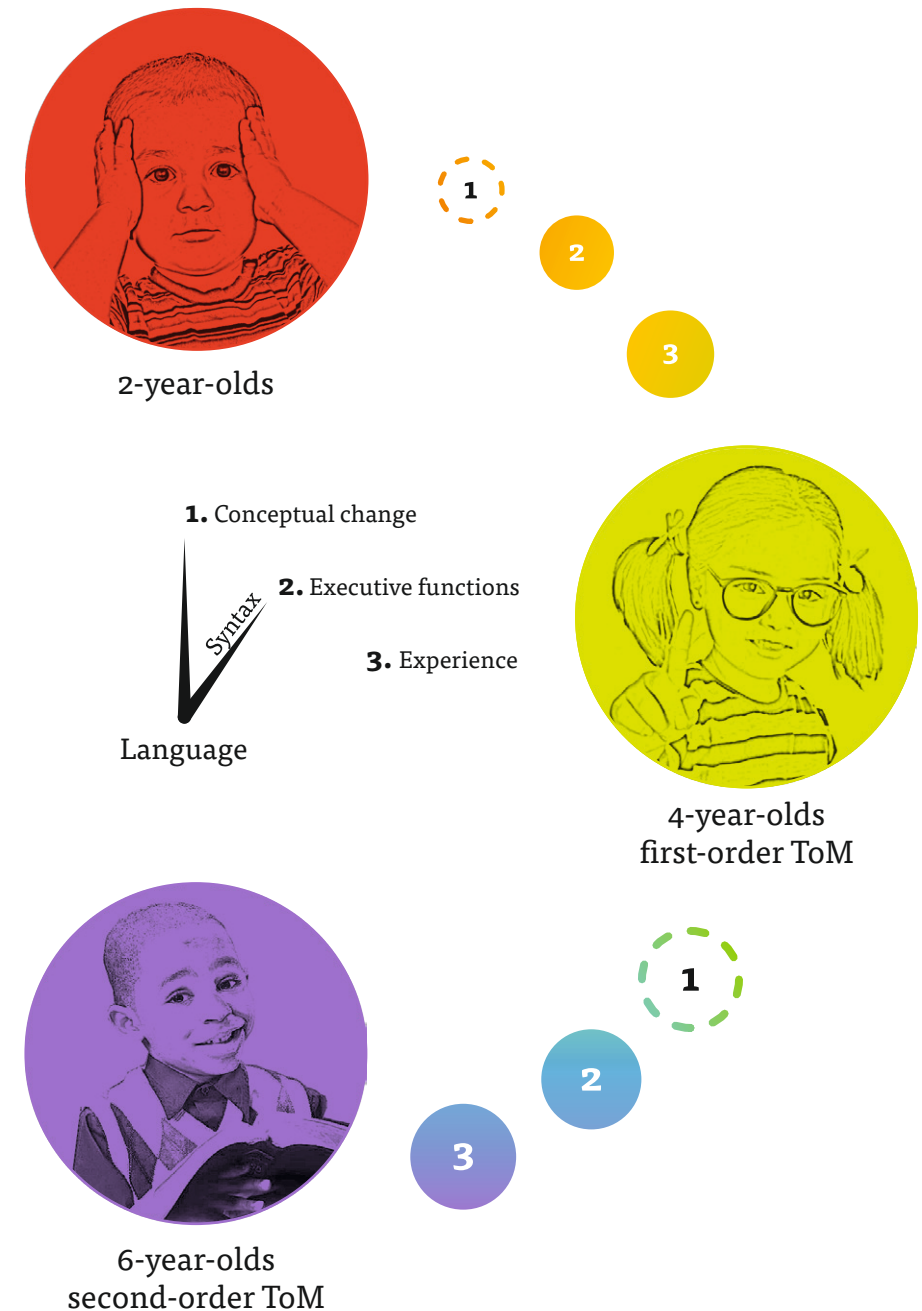
Zoals we echter hebben laten zien in Hoofdstuk 6, is werkgeheugen alleen niet genoeg. Kinderen hebben ook cognitieve controle nodig om een juist antwoord te geven. Voor eerste-orde theory of mind redeneren is het belangrijk om kinderen feedback te geven met een verklaring, zodat ze begrijpen dat cognitieve controle nodig is. Voor tweede-orde theory of mind redeneren is het daarentegen belangrijk dat kinderen voorbeelden zien waarmee ze hun incorrecte strategie kunnen corrigeren naar een tweede-orde theory of mind strategie.

Conclusies

In dit proefschrift hebben we computationele cognitieve modellen gecombineerd met empirisch onderzoek om de ontwikkeling van tweede-orde theory of mind in kinderen te onderzoeken. Na eerdere theorieën te hebben bestudeerd hebben we op basis van deze theorieën computationele cognitieve modellen gebouwd. Het bouwen van deze modellen stelde ons ertoe in staat om cognitieve modellen en cognitieve concepten precies te implementeren in plaats van deze concepten te gebruiken zonder onderscheid te maken in hun onderliggende componenten. De simulatieresultaten van de modellen brachten nieuwe voorspellingen aan het licht die we empirisch hebben getest. Bovendien hebben we de computationele modelleeraanpak gebruikt om mechanismen te begrijpen die ten grondslag lagen aan beschikbare data in de literatuur. Op basis van onze modellen en empirische resultaten en samen met de bestaande literatuur, zou ik een chronologie willen voorstellen voor de factoren die bijdragen aan de ontwikkeling van tweede-orde theory of mind in kinderen, die is weergegeven in Figuur 7.1

De ontwikkeling van eerste-orde theory of mind is een noodzakelijke voorwaarde voor de ontwikkeling van tweede-orde theory of mind. Het lijkt waarschijnlijker dat er een *nieuw inzicht* nodig is voor de ontwikkeling van eerste-orde theory of mind dan voor de ontwikkeling van tweede-orde theory of mind. Dit komt doordat jonge kinderen eerst het inzicht moeten krijgen dat anderen mentale toestanden kunnen hebben (bijv. overtuigingen, intenties) die mogelijk anders zijn dan hun eigen mentale toestanden. Dit inzicht lijkt zich te ontwikkelen met behulp van pragmatiek en semantische componenten van de taal.

Na een mogelijk nieuw inzicht kampen jonge kinderen nog steeds met de complexiteit van eerste-orde false-belieftaken op het gebied van executive functies zoals werkgeheugen en cognitieve controle. Ze hebben efficiënte redeneerregels nodig om te representeren dat de overtuigingen van een ander incorrect zijn (d.w.z., van nulde-orde theory of mind redeneren naar eerste-orde theory of



Figuur 7.1. Een chronologie voor de factoren die bijdragen aan de ontwikkeling van tweede-orde theory of mind in kinderen.

mind redeneren), en moeten daarom een mogelijke *serial processing bottleneck* van werkgeheugen overwinnen. Taken in het dagelijks leven die gebruik maken van werkgeheugenstrategieën en cognitieve coördinatie dragen bij aan het verkrijgen van efficiënte redeneerregels (Hoofdstuk 5 en Hoofdstuk 6). Daarnaast helpt de syntactische component van taal bij het segmenteren van informatie in chunks die door een bottleneck in het werkgeheugen kunnen komen. Tenslotte hebben kinderen ervaring en feedback (verbaal of non-verbaal) nodig om te begrijpen dat een eerste-orde theory of mind redeneerstrategie nodig is voor een bepaalde taak of situatie.

Nadat ze eerste-orde theory of mind hebben geleerd, moeten kinderen opnieuw de complexiteit van tweede-orde theory of mind overwinnen op het gebied van taak en werkgeheugenstrategieën. Syntactische recursie in het taaldoel is een geschikte manier om ze te helpen om efficiënte redeneerregels te ontwikkelen en om hiërarchisch ingebedde mentale toestanden op een seriële manier te verwerken ten behoeve van de *serial processing bottleneck* in het werkgeheugen (Hoofdstuk 4).

Kinderen die kunnen omgaan met complexe werkgeheugenstrategieën hebben nog steeds ervaring nodig met tweede-orde theory of mind redeneren om hun incorrecte eerste-orde redeneerstrategie aan te passen tot de correcte tweede-orde redeneerstrategie (Hoofdstuk 2 en Hoofdstuk 3). Deze ervaring kan worden verkregen door het lezen van verhalen en het spelen van spelletjes waarbij tweede-orde theory of mind nodig is, maar ook door sociale interacties met familie en vrienden.

References

A

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30–60.
- Agresti A. (2013). *Categorical Data Analysis* (third ed.). New York: John Wiley & Sons.
- Aksu-Koç A. (1988). *The Acquisition of Aspect and Modality: The Case of Past Reference in Turkish*. Cambridge: Cambridge University Press.
- Anderson, J. R. & Lebiere, C. J. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.
- Anderson, J. R. & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving and F. I. M. Craik (Eds.), *Handbook of Memory* (pp. 557–570). New York: Oxford University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review*, 111, 1036–1060.
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.
- Apperly I. A. & Butterfill S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–970.
- Appleton, M. & Reddy, V. (1996). Teaching three year-olds to pass false belief tests: A conversational approach. *Social Development*, 5, 275–291.
- Arslan, B., Taatgen, N., & Verbrugge, R. (2013). Modeling developmental transitions in reasoning about false beliefs of others. In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling* (pp. 77–82). Ottawa: Carleton University.
- Arslan, B., Verbrugge, R., Taatgen, N. & Hollebrandse, B. (2014). Teaching children to attribute second-order false belief: A training study. In J. Szymanik & R. Verbrugge (Eds.), *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives* (Vol. 1208, pp. 1–5). Groningen: CEUR Workshops Proceedings.
- Arslan, B., Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2015a). Teaching children to attribute second-order false beliefs: A training study with feedback. In D. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 108–113). Austin, TX: Cognitive Science Society.
- Arslan, B., Wierda, S., Taatgen, N., & Verbrugge, R. (2015b). The role of simple and complex working memory strategies in the development of first-order false belief reasoning: A computational model of transfer of skills. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.), *Proceedings of the 13th International Conference on Cognitive Modeling* (pp. 100–105). Groningen: University of Groningen.
- Arslan, B., Taatgen, N. A., & Verbrugge, R. (2017a). Five-year-olds' systematic errors in second-order false belief tasks are due to first-order theory of mind strategy selection: A computational modeling study. *Frontiers in Psychology*, 8:275. doi: 10.3389/fpsyg.2017.00275
- Arslan, B., Hohenberger, A., & Verbrugge, R. (2017b). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PLOS ONE* 12(1): e0169510. doi: 10.1371/journal.pone.0169510

- Arslan, B., Verbrugge, R., & Taatgen, N. (2017c). Cognitive control explains the mutual transfer between dimensional change card sorting and first-order false belief understanding: A computational modeling study on transfer of skills. *Journal of Biologically Inspired Cognitive Architectures*. <http://dx.doi.org/10.1016/j.bica.2017.03.001>
- Astington J. W., Pelletier J., & Homer B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology*, 20, 131-144.
- Astington, J. W. & Baird, J. (2005). *Why Language Matters for Theory of Mind*. Oxford: Oxford University Press.

B

- Baillargeon, R., Scott, R. M., He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science*, 14, 110-118.
- Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., Wu, D., & Bian, L. (2015). Psychological and sociomoral reasoning in infancy. In M. Mikulincer, P. R. Shaver (Eds.), E. Borgida, & J. A. Bargh (Assoc. eds.), *APA Handbook of Personality and Social Psychology: Vol. 1. Attitudes and Social Cognition* (pp. 79-150). Washington, DC: American Psychological Association.
- Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking and Reasoning*, 15, 337-354.
- Bello, P. & Cassimatis, N. (2006). Developmental accounts of theory-of-mind acquisition: Achieving clarity via computational cognitive modeling. In R. Sun & N. Miyake (Eds.), *Proceeding of the 28th Annual Meeting of the Cognitive Science Society* (pp. 1014-1019). Vancouver: Cognitive Science Society.
- Belpaeme, T., Baxter, P., de Greeff, J., Kennedy, J., Read, R., Looije, R. et al. (2013). Child-robot interaction: Perspectives and challenges. In: G. Herrmann, M.J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Social Robotics* (Vol. 8239, pp. 452-459). New York, NY: Springer.
- Benson, J. E., Sabbagh, M. A., Carlson, S. M. & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology*, 49(9), 1615-1627.
- Bianco, F., Lecce, S., & Banerjee, R. (2016). Conversations about mental states and theory of mind development during middle childhood: A training study. *Journal of Experimental Child Psychology*, 149, 41-61.
- Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Science*, 8 (6), 255-260.
- Birch, S. A. J., & Bloom, P. (2007) The curse of knowledge in reasoning about false belief. *Psychological Science*, 18, 382-386.
- Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 36(2), 363-382.
- Bosco, F. M. & Gabbatore, I. (2017). Sincere, deceitful, and ironic communicative acts and the role of the theory of mind in childhood. *Frontiers in Psychology*, 8(21), doi: 10.3389/fpsyg.2017.00021.

- Braüner, T., Blackburn, P., & Polyanskaya, I. (2016). Second-order false-belief tasks: Analysis and formalization. In *Proceedings of 23rd International Workshop on Logic, Language, Information, and Computation* (pp. 125-144). Springer Berlin Heidelberg.
- Bull, R., Phillips, L. H., & Conway, C. (2008). The role of control functions in mentalizing: Dual-task studies of Theory of Mind and executive function. *Cognition*, 107, 663-672.
- Buss, A., & Spencer, J.P. (2008). The emergence of rule-use: A dynamic neural field model of the DCCS. In B.C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 463-468). Austin, TX: Cognitive Science Society.

C

- Caillies, S. & Le Sourn-Bissaoui, S. (2013). Nondecomposable idiom understanding in children: Recursive theory of mind and working memory. *Canadian Journal of Experimental Psychology*, 67, 108-16.
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *The Journal of Political Economy*, 97 (5), 1232- 1254.
- Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3), 861-898.
- Carlson, M. S., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, 69, 672-691.
- Carlson, S. M. & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72, 1032-1053.
- Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant Child Development*, 11, 73-92.
- Carlson, S.M., Moses, L.J., & Claxton, L.J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, 87, 299-319.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, 28(2), 595-616.
- Carlson, S. M., Claxton, L. J., & Moses, L.J. (2014). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development*, 16 (1), 186-197.
- Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, 28, 141-172.
- Carruthers, P. (2015). Mindreading in adults: Evaluating two-systems views. *Synthese*, 192, 1-16.
- Carruthers, P. (2016a). Two systems for mindreading? *Review of Philosophy and Psychology*, 7, 141-162.
- Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17, 193-199
- Clements, W., Rustin, C. L., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: A training study of false belief. *Developmental Science*, 3, 81-92.
- Cohen, J. D., Servan-Schreiber D., & McClelland J. L. (1992). A parallel distributed processing approach to automaticity. *The American Journal of Psychology*, 105(2), 239-269.

- Cole, K. & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology*, 18, 279–295.
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786.
- Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., et al. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, 132(1), 113–32.

D

- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Davis, H. L. & Pratt C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology*, 47 (1), 25–31.
- de Villiers, J. G & de Villiers, P. A. (2000). Linguistic determinism and false belief. In: P. Mitchell & K. Riggs (Eds.), *Children's Reasoning and the Mind* (pp. 191–228). Hove, England: Psychology Press.
- de Villiers, J. G., & Pyers, J. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development*, 17, 1037–1061.
- de Villiers, J. G. (2005). Can language acquisition give children a point of view? In: Astington, J. & Baird, J. (Eds). *Why Language Matters for Theory of Mind* (pp. 186–219). Oxford: Oxford University Press.
- de Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua*, 117, 1858–1878.
- de Villiers, J. G. & de Villiers, P. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, 34, 313–328.
- de Villiers, J., Hobbs, K., & Hollebrandse, B. (2014). Recursive complements and propositional attitudes. In M. Speas & T. Roeper (Eds.), *Recursion: Complexity in Cognition*. Theoretical Studies in Psycholinguistics, (Vol. 43, pp. 221–242). Heidelberg, New York: Springer.
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 4, 568–570.
- Devine, R. T. & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development*, 84, 989–1003.
- Devine, R.T. & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, 85, 1777–1794.
- Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, 38, 352–362.
- Doebel, S. & Zelazo, P. D. (2015). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, 38, 241–268.
- Dumontheil, I., Apperly, I. A., & Blakemore, S. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331–338.

E-F

- Elio, R. (1986). Representation of similar well-learned cognitive procedures. *Cognitive Science*, 10, 41–73.
- Engle, R., Tuholski, S. W., Laughlin, J. E., & Conway, A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 130, 169–183.
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40, 760–768.
- Etel, E. & Yagmurlu, B. (2015). Social competence, theory of mind, and executive function in institution-reared Turkish children. *International Journal of Behavioral Development*, 39(6), 519–529.
- Filippova, E. & Astington, J. W. (2008) Further development in social reasoning revealed in discourse irony understanding. *Child Development*, 79, 126–38.
- Fischhoff, B. (1975). Hindsight does not equal foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2008). Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, 17(4), 417–442.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology*, 50(8), 2085–2092.

G

- Girli, A. & Tekin, D. (2010). Investigating false belief levels of typically developed children and children with autism. *Procedia – Social and Behavioral Sciences*, 2(2), 1944–1950.
- Gleitman, L. & Papafragou, A. (2005). Language and thought. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 633–661). Cambridge: Cambridge University Press.
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
- Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., et al. (2006). Intuitive theories of mind: A rational approach to false belief. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1382–1390). Vancouver: Cognitive Science Society.
- Goodie, A.S., Doshi, P., & Young, D. L. (2012) Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, 25(1), 95–108.
- Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257–293). New York, NY: Cambridge University Press.
- Gordon, A. C. & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology*, 68, 70–83.

H

- Hale, C. M. & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, 6, 346–359.
- Hansen M. B. (2010). If you know something, say something: Young children's problem with false beliefs. *Frontiers in Psychology*, 1(23), 1–7.
- Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129–154.
- Harris, P.L. (2002). From simulation to folk psychology: The case for development. *Mind and Language*, 7, 120–144.
- Hasselhorn, M., Mähler, C., & Grube D. (2005). Theory of mind, working memory, and verbal ability in preschool children: The proposal of a relay race model of the developmental dependencies. In: W. Schneider, R. Schumann-Hengsteler, & B. Sodian, (Eds.), *Young Children's Cognitive Development: Interrelations Among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind* (pp. 219–237). Mahwah, NJ: Erlbaum,
- Hedden, T. & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, 85(1), 1–36.
- Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, 18, 167–170.
- Hendriks, P., van Rijn, H., & Valkenier, B. (2006). Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua*, 117(11), 1879–1896.
- Henning, A., Spinath, F. M., & Aschersleben, G. (2011). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology*, 108, 513–531.
- Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, 9, 233–247.
- Heyes, C. M. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, 9(2), 131–143.
- Heyes, C. M. & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, 344(6190), 1243091.
- Hiatt, L. M. & Trafton, J. G. (2010). A cognitive model of theory of mind. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 91–96). Philadelphia, PA: Drexel University.
- Hiatt, L. M., & Trafton, J. G. (2015). Understanding second-order theory of mind. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (pp. 167–168). Portland, OR: ACM.
- Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., Curtiss, J., & Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition*, 150, 200–212.
- Hollebrandse, B., Hobbs, K., de Villiers, J., & Roeper, T. (2008). Second order embedding and second order false belief. In A. Gavarró & M. J. Freitas (Eds.), *Language Acquisition and Development*. Proceedings of GALA 2007 (pp. 270–280). Cambridge Scholar Press.

- Hollebrandse, B., Van Hout, A. & Hendriks P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, 191, 321–333.
- Hollebrandse, B. & Roeper, T. (2014). Empirical results and formal approaches to recursion in acquisition. In: M. Speas & T. Roeper (Eds.), *Recursion: Complexity in Cognition Theoretical Studies in Psycholinguistics* (pp. 179–219). Heidelberg, NY: Springer
- Holmes, H. A., Black, C. & Miller, S. A. (1996). A cross-task comparison of false belief understanding in a Head-Start population. *Journal of Experimental Child Psychology*, 63, 263–285.
- Hsu, Y. K. & Cheung, H. (2013). Two mentalizing capacities and the understanding of two types of lie telling in children. *Developmental Psychology*, 49, 1650–1659.
- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16, 233–253.
- Hughes, C. (2002). Executive functions and development: Emerging themes. *Infant and Child Development*, 11, 201–209.

K

- Kahneman D. (1973). *Attention and Effort*. Englewood Cliffs, N.J: Prentice-Hall.
- Keenan, T., Olson, D. R., & Marini, Z. (1998). Working memory and children's developing understanding of mind. *Australian Journal of Psychology*, 50, 76–82.
- Keysar, B., Barr, D.J., Balin, J.A., & Brauner, J.S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32–38.
- Keysar, B., Lin, S., & Barr, D.J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25–41.
- Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89(2), 191–204.
- Kloo, D. & Perner J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development*, 74(6), 1823–1839.
- Kloo, D. & Perner, J. (2008). Training theory of mind and executive control: A tool for improving school achievement? *Mind, Brain, and Education*, 2, 122–127.
- Kovács, Á. M., Téglás, E., Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830–1834.
- Kuijper, S. (2016). *Communication Abilities of Children with ASD and ADHD*, PhD thesis, University of Groningen. 2016.
Available: [http://www.rug.nl/research/portal/en/publications/communication-abilities-of-children-with-asd-and-adhd\(d2709b53-6233-4c60-86d4-d72d5eefa9be\).html](http://www.rug.nl/research/portal/en/publications/communication-abilities-of-children-with-asd-and-adhd(d2709b53-6233-4c60-86d4-d72d5eefa9be).html)

L

- Laird, J. E. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.
- Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014a). Training preschoolers on first-order false belief understanding: Transfer on advanced ToM skills and metamemory. *Child Development*, 85, 2404–2418.

- Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2014b). Promoting theory of mind during childhood: A training program. *Journal of Experimental Psychology*, 126, 52–67.
- Leslie A. M. (1987). Pretence and representation: The origins of “theory of mind.” *Psychological Review*, 94, 412–426.
- Leslie, A. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the Mind* (pp. 119–148). New York: Cambridge University Press.
- Leslie, A. M. & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, 1, 247–254.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind’. *Trends in Cognitive Sciences*, 8 (12), 528–533.
- Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, 50, 45–85.
- Lewis, S., Hacquard, V. & Lidz, J. (2012). The semantics and pragmatics of belief reports in preschoolers. *Proceedings of SALT 22* (pp. 247–267).
- Ling, D. S., Wong, C. D., & Diamond, A. (2015). Do children need reminders on the day–night task, or simply some way to prevent them from responding too quickly? *Cognitive Development*, 37, 1–6.
- Lockl, K. & Schneider, W. (2007). Knowledge about the mind: Links between theory of mind and later metamemory. *Child Development*, 78, 148–167.
- Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, 10(3), 184–189.
- Lohmann, H. & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, 74, 1130–1144.

M

- Marcovitch, S., & Zelazo, P.D. (2000). A generative connectionist model of the development of rule use in children. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 334–339). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marr, D. (1982). *Vision*. Freeman Publishers.
- McKinnon, M.C. & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, 102, 179–218.
- Meijering, B., Van Maanen, I., Van Rijn, H., & Verbrugge, R. (2010). The facilitative effect of context on second-order social reasoning. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1423–1428). Austin, TX: Cognitive Science Society.
- Meijering, B., Taatgen, N. A., Van Rijn, H., & Verbrugge, R. (2014). Modeling inference of mental states: As simple as possible, as complex as necessary. *Interaction Studies*, 15, 455–477.
- Melot, A. N. & Angeard, N. (2003). Theory of mind: Is training contagious? *Developmental Science*, 6, 178–184.
- Meyer, D. E. & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104(1), 3–65.

- Meyer, D. E. & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, 104(4), 749–791.
- Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, 135, 749–773.
- Miller, S. A. (2012). *Theory of Mind: Beyond the Preschool Years*. Hove: Psychology Press.
- Milligan, K., Astington, J. W., & Dack, L. (2007). Language and theory of mind: A meta analysis of the relation between language ability and false belief understanding. *Child Development*, 78, 622–646.
- Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of reality bias in adults but not children. *Cognition*, 59, 1–21.
- Mitchell, P. & Taylor, L. M. (1999) Shape constancy and theory of mind: Is there a link? *Cognition*, 70, 167–190.
- Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: a neural network model of perseveration, dissociation and decalage. *Developmental Psychobiology*, 40, 255–265.
- Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2005). On the specificity of the relation between executive function and children's theories of mind. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young Children's Cognitive Development: Interrelations Among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind* (pp. 131–145). Mahwah, NJ: Erlbaum.
- Müller, U., Zelazo, D. P., & Imrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development*, 20, 173–189.

N-O

- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 283–308). San Diego, CA: Academic Press.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258.
- Özge, D., Marinis, T., & Zeyrek D. (2010). Comprehension of subject and object relative clauses in monolingual Turkish children. In: S. Ay, Ö. Aydın, İ. Ergenç, S. Gökmen, S. İşsever, & D. Peçenek (Eds.), *Proceedings of the Fourteenth International Conference of Turkish Linguistics* (pp. 341–351). Wiesbaden: Harrasowitz Verlag.
- Özoran D. (2009). *Cognitive Development of Turkish Children on the Relation of Evidentiality and Theory of Mind*. Middle East Technical University.

P

- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York: Cambridge University Press.
- Penn, D. C. & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind’. *Philosophical Transactions of the Royal Society B*, 362(1480), 731–744.

- Penner, Z. (1999). *Screeningverfahren zur Feststellung von Störungen in der Grammatikentwicklung*. Luzern, Switzerland: Edition SZH/SPC.
- Pérez-Leroux, A. T., Castilla-Earls A. P., Bejar, S., Massam, D. (2012). Elmo's sister's ball: The problem of acquiring nominal recursion. *Language Acquisition*, 19(4), 301–311.
- Perner, J. & Wimmer, H. (1985). "John thinks that Mary thinks that...": Attribution of second-order beliefs by 5- to 10-year old children. *Journal of Experimental Child Psychology*, 5, 125–137.
- Perner, J. (1998). The meta-intentional nature of executive functions and theory of mind. In P. Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary Themes* (pp. 270–283). Cambridge, England: Cambridge University Press.
- Perner, J. & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, 3, 337–344.
- Perner, J., Kain, W., & Barchfeld, P. (2002). Executive control and higher-order theory of mind in children at risk of ADHD. *Infant and Child Development*, 11(2), 141–158.
- Perner, J. & Lang, B. (2002). What causes 3-year-olds' difficulty on the Dimensional Change Card Sorting task? *Infant and Child Development*, 11, 93–105.
- Perner, J., & Ruffman, T. (2005). Infant's insight into the mind: How deep? *Science*, 308, 214–216.
- Peterson, C. C. & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language*, 15, 123–145.
- Piaget, J. (1965). *The Child's Conception of Number*. New York: Norton.
- Pratt, C. & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, 61, 973–983.
- Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526.

R

- Raijmakers, M. E. J., Mandell, D. J., Van Es, S. E., & Counihan, M. (2014). Children's strategy use when playing strategic games. *Synthese*, 191(3), 355–370.
- Rieskamp, J. & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.
- Roeper, T. (2007). *The Prism of Grammar: How Child Language Illuminates Humanism*. Cambridge, MA: MIT Press.
- Rosenbloom, P. S. (2013). The Sigma cognitive architecture and system. *AISB Quarterly*, 136, 4–13.
- Rubio-Fernández, P. (2017). Can we forget what we know in a false-belief task? An investigation of the true-belief default. *Cognitive Science*, 41, 218–241.
- Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, 34, 161–174.
- Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, 73(3), 734–751.

- Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development*, 18, 139–158.
- Russell, J. (1996). *Agency: Its Role in Mental Development*. Hove: Erlbaum.
- ## S
- Sabbagh, M. A., Hopkins, S. F. R., Benson, J. E., & Flanagan, R. J. (2010). Conceptual change and preschoolers' theory of mind: Evidence from load–force adaptation. *Neural Networks*, 23, 1043–1050.
- Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multi-tasking. *Psychological Review*, 115, 101–130.
- Scholl, B., & Leslie, A. (1999). Modularity, development and "theory of mind". *Mind & Language*, 14, 131–153.
- Schmidt, A. M. G. & Westendorp, F. (2011). *Jip en Janneke*. Amsterdam: Querido.
- Shuliang, M., Yanjie, S., Sabbagh, M. A., & Jiaming, X. (2014). Sentential complements and false belief understanding in Chinese Mandarin-speaking preschoolers: A training study. *Cognitive Development*, 29, 50–61.
- Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *PNAS*: 1609203113v1-201609203.
- Shah A. (2012). Psychological and neuroscientific connections with reinforcement learning. In M. Wiering & M. Van Otterlo (Eds.), *Reinforcement Learning: State of the Art* (pp. 507–537). Verlag: Springer.
- Singley, M. K., & Anderson, J. R. (1985). The transfer of text-editing skill. *International Journal of Man-Machine Studies*, 22, 403–423.
- Slade, L. & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, 23, 117–141.
- Slaughter, V. & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, 67 (6), 2967–2988.
- Smith, M., Apperly, I. & White, V. (2003). False belief reasoning and the acquisition of relative clause sentences. *Child Deveopment*, 74, 1709–1719.
- Sodian B., Taylor C., Harris P., & Perner J. (1992). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development*, 62, 468–483.
- Stenning K. & van Lambalgen M. (2008). *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press.
- Stevens, C. A., Taatgen, N. A., Cnossen, F. (2016). Instance-based models of metacognition in the Prisoner's Dilemma. *Topics in Cognitive Science*, 1–13.
- Stiller, J. & Dunbar, R.I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, 29(1), 93–104.
- Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30(3), 395–402.
- Sutton, R. S., and Barto, A. G. (1998) *Reinforcement Learning: An Introduction*. MIT Press.

T

- Taatgen, N. A., & Anderson, J. (2002). Why do children learn to say “broke”? A model of learning the past tense without feedback. *Cognition*, 86, 123-155.
- Taatgen, N. A. & Lee, F. J. (2003). Production compilation: A simple mechanism to model complex skill acquisition. *Human Factors*, 45: 61–76.
- Taatgen, N. A., Lebiere, C. & Anderson, J. R. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (pp. 29-52). Cambridge, UK: Cambridge University Press.
- Taatgen, N.A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, 120(3), 439–471.
- Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders*, 24, 577-586.
- Talwar, V. & Lee, K. (2008). Social and cognitive correlates of children's lying behaviour. *Child Development*, 79 (4), 866-881.
- Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39 (2), 195-217.
- Triona, L. M., Masnick, A. M., & Morris, B. J. (2002). What does it take to pass the false belief task? An ACT-R model. In D. G. Wayne & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 1045). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

U-V

- Unsworth, N. & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54(1), 68–80.
- Ünal G. (2008). *Release from Proactive Interference and its Relations to Executive Functions: A Developmental Study on Turkish Children* [Internet]. Middle East Technical University. Available: <https://etd.lib.metu.edu.tr/upload/12609984/index.pdf>
- Van Bers, B. M. C. W., Visser, I., van Schijnder T. J. P., Mandell, D. J., & Raijmakers, E. J. (2011). The dynamics of development on the Dimensional Change Card Sorting task. *Developmental Science*, 14(5), 960–971.
- Van Rijn, H., Van Someren, M., & Van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, 27(2), 227-257.
- Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, 37, 731–766.
- Verbrugge, R. (2009). Logic and social cognition: The facts matter, and so do computational models. *Journal of Philosophical Logic*, 38, 649–680.

W

- Wahl, S. & Spada, H. (2000). Children's reasoning about intentions, beliefs and behaviour. *Cognitive Science Quarterly*, 1, 5–34.
- Walker, R. F. & Murachver, T. (2012). Representation and theory of mind development. *Developmental Psychology*, 48(2), 509–20.

- Watanabe, K., & Funahashi, S. (2015). A dual-task paradigm for behavioral and neurobiological studies in nonhuman primates. *Journal of Neuroscience Methods*, 246, 1–12.
- de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, 199–200, 67–92.
- de Weerd, H., Verbrugge, R., & Verheij, B. (2015). Higher-order theory of mind in the Tacit Communication Game. *Biologically Inspired Cognitive Architectures*, 11, 10–21.
- de Weerd, H., Verbrugge, R., & Verheij, B. (2017). Negotiating with other minds: The role of recursive theory of mind in negotiation with incomplete information. *Autonomous Agents and Multi-Agent Systems*, 31(2), 250-287. doi:10.1007/s10458-015-9317-1.
- Wellman, H. M. (1985). The child's theory of mind: The development of conceptions of cognition. In S. R. Yussen (Ed.), *The Growth of Reflection in Children* (pp. 169–206). San Diego, CA: Academic Press.
- Wellman, H.M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
- Wellman, H. M., & Gopnik, A. (1994). The “theory theory.” In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257–293). New York: Cambridge University Press.
- Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655–684.
- Westra, E. & Carruthers, P. (2017). Pragmatic development explains the theory-of-mind scale. *Cognition*, 158, 165–176.
- Wierda, S. & Arslan, B. (2014). Modeling theory of mind in Actransfer. In J. Szymanik & R. Verbrugge (Eds.), *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives* (Vol. 1208, pp. 40-44). Groningen: CEUR Workshops Proceedings.
- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Wright, J. R. & Leyton-Brown, K. (2010). Beyond equilibrium: Predicting human behavior in normal-form games. In *Proceedings of the 24th Conference on Artificial Intelligence* (pp. 901-907).

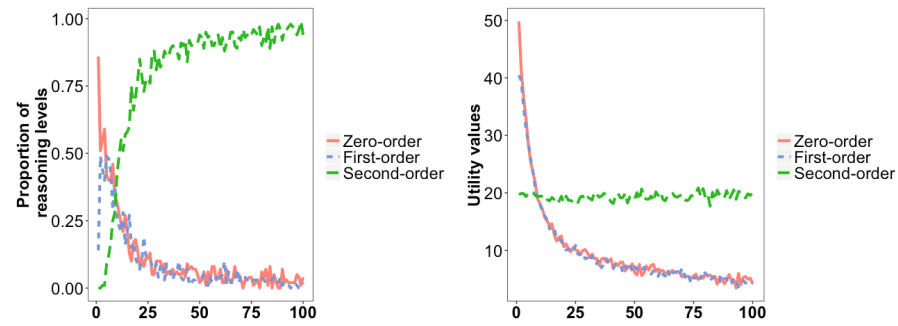
Z

- Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68 (Serial No. 274), i-151.
- Zhang, L. J., Wu, N., & Zheng, Y. (2008). The training effects of different linguistic interference on young Chinese performance on a theory of mind task. *Acta Psychologica Sinica*, 40(7), 819-827

S1 Materials

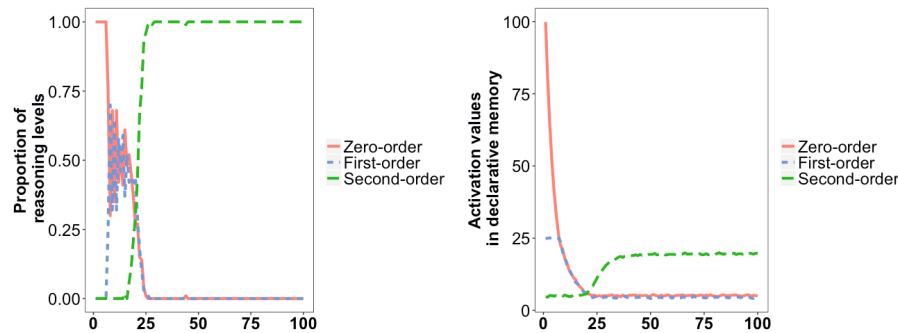
Reinforcement learning model results with different utility values

In order to show that the initial utility values do not change the qualitative predictions of the model, the utility values are set to 50, 40, 20 for the zero-order, first-order and second-order ToM strategies respectively. Rest of the parameters and the number of repetitions are kept the same with the original model.



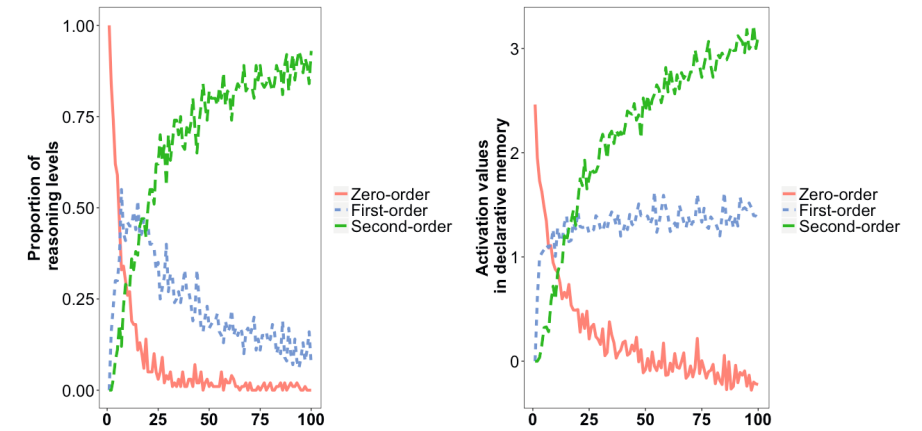
Reinforcement learning model results with a lower noise value

In order to show that the noise values do not change the qualitative predictions of the model, the noise value is set to 1 (it is set to 3 in the original model). Rest of the parameters and the number of repetitions are kept the same as the original model.



Instance-based learning model results with a higher noise value

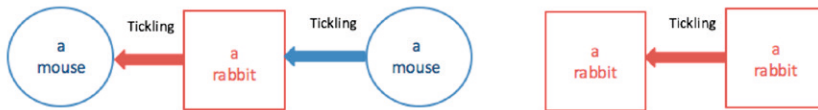
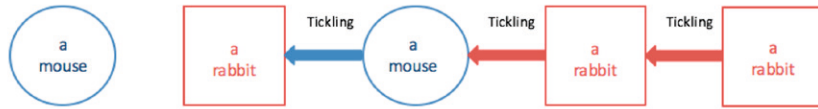
In order to show that the noise values do not change the qualitative predictions of the model, the noise value is set to 0.5 (it is set to 0.1 in the original model). Rest of the parameters and the number of repetitions are kept the same as with the original model.



S2 Materials

Second-order relative clause task (REL_2) questions and schematic figures

REL_2 Practice Question: In which picture is there a rabbit that is tickling a mouse that is tickling a rabbit? (“Hangi resimde tavşanı gıdıklayan fareyi gıdıklayan bir tavşan var?”)



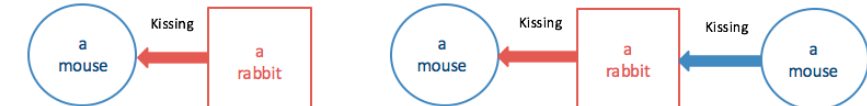
REL_2 Lion-Gorilla: In which picture is there a lion that is biting a lion that is biting a gorilla? (“Hangi resimde gorili ısırın aslanı ısırın bir aslan var?”)



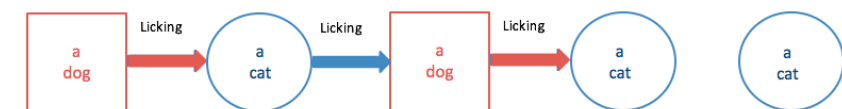
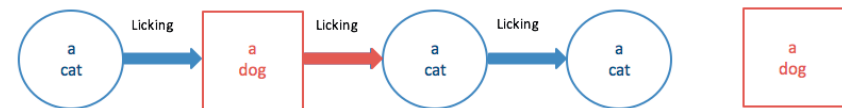
REL_2 Horse-Camel: In which picture is there a horse that is caressing a camel that is caressing a horse? (“Hangi resimde atı okşayan deveyi okşayan bir at var?”)



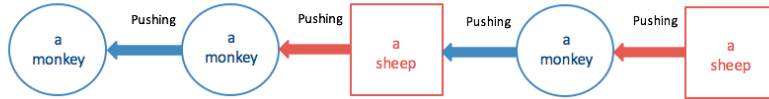
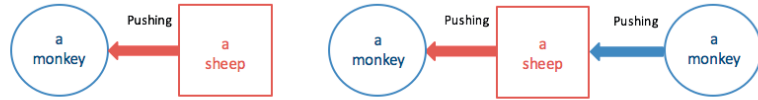
REL_2 Mouse-Rabbit: In which picture is there a mouse that is kissing a rabbit that is kissing a mouse? (“Hangi resimde fareyi öpen tavşanı öpen bir fare var?”)



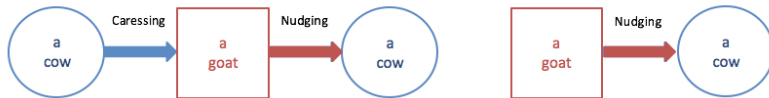
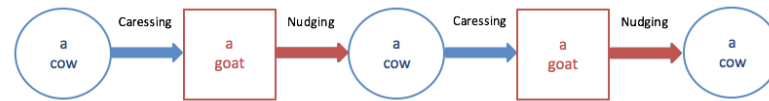
REL_2 Cat-Dog: In which picture is there a dog that is licking a cat that is licking a dog? (“Hangi resimde köpeği yalayan kediye yalan bir kopek var?”)



REL_2 Sheep-Monkey: In which picture is there a sheep that is pushing a monkey that is pushing a sheep? (“Hangi resimde kuzuyu iten maymunu iten bir kuzu var?”)



REL_2 Goat-Cow: In which picture is there a goat that is nudging a cow that is caressing a goat? (“Hangi resimde keçiyi okşayan ineği boynuzlayan bir keçi var?”)



Word span task stimuli (WST)

SETS OF 2

Köşk - Muz

Pil - Üst

Buz - Dört

SETS OF 3

Göl - Saç - Tuz

Sev - Kürk - Bel

Kir - Ut - Pas

SETS OF 4

Kaş - Sos - Gök - Yat

Cam - But - Sal - Köy

Zar - Kuş - Tüm - Can

SETS OF 5

Suc - Kek - Böl - Top - Zam

Bal - Kurt - As - Tat - Cöp

Ot - Son - Türk - Seç - Kol

SETS OF 6

Hak - Sus - Tek - Mum - Dip - Kar

Kes - Bin - Ter - Aşk - Yut - Sel

Tren - Kel - Söz - An - Koy - Tez

SETS OF 7

Ak - Top - Su - Alt - Bey - Bol - Mart

Tel - Poz - At - Bil - Yok - Fes - Tür

Kış - Ver - Han - Bot - Yıl - Post - Kül

SETS OF 8

Tam - Bak - Uç - Göz - Hal - Boş - Ek - Yurt

Üç - Kas - Al - Mülk - Bir - Tut - Dil - Kum

Bul - Pek - On - Fal - Var - El - Ses - Genç

Listening span task stimuli (LST)

Used in the trials

1. Çocuklar okula gider.
2. Balıklar havada yaşar.
3. Ağaçlar dans eder.

SETS OF 2

1. Biber acıdır.
2. Kediler okulda çalışır.
2. Filler çok küçüktür.
2. Ayakkabı ayağa giyilir.
3. İnsanlar saçlıdır.
2. Çiçekler fare kovalar.
4. 1. Ayılar araba sürer.
2. Havuçlar turuncudur.
5. 1. Gece karanlıktır.
2. Portakallar suda yaşar.
6. 1. Ateş sıcaktır.
2. Balıklar konusur.

SETS OF 3

1. Otobüslerle tatile gideriz.
2. Toplar karedir.
3. Öğretmenler ağaçta yetişir.
2. Muzlar bisiklete biner.
2. Elimiz beş parmaklıdır.
3. Soğan acıdır.
3. 1. Otobüsler oyuncakla oynar.
2. Kuşlar kanatlıdır.
3. Elmalar ağaçta yetişir.
4. 4. Piyanolar müzik çalar.
5. Kardeşlerimiz kuyrukludur.
6. Burnumuzla görürüz.
5. 4. Ayağımız çenelidir.
5. Güneş sıcaktır.
6. Taşlar serttir.
6. 4. Kaşıklarla yazı yazarız.
5. Limon şarıdır.
6. Köpekler kedileri kovalar.

SETS OF 4

1. 1. Zürafalar uzun boyludur.
2. Çiçekler pasta sever.
3. Portakallar kulaklıdır.
4. Öğretmenler okulda çalışır.
2. 1. Otobüsler konusur.
2. Bankalardan para çekimiz.
3. Kışlar sıcaktır.
4. Pastalar tatlıdır.
3. 1. Gökyüzü kırmızıdır.
2. Bebekler ağlar.
3. Köpekler konusur.
4. Muzlar tatlıdır.
4. 1. Armutlar mavidir.
2. Şapkalar başa giyilir.
3. Tavşanlar saati gösterir.
4. Filler büyüktür.
5. 1. İnsanlar iki ayaklıdır.
2. Portakallar siyahtır.
3. Kediler futbol oynar.
4. Kitapları okuruz.
6. 1. Tavşanlar ağaçta yetişir.
2. Biberler yeşildir.
3. Portakallar markette satılır.
4. İnsanlar üç gözlüdür.

SETS OF 5

1. 1. Babalar kanatlıdır.
2. Dondurma soğuktur.
3. Portakallar gitar çalar.
4. Arabalar benzinle çalışır.
5. Fareler çok büyüktür.
2. 1. Havuçlar mavidir.
2. Kulaklarımızla görürüz.
3. Portakallar turuncudur.
4. Tavuklar yumurta yapar.
5. Bıçak keskindir.
3. 1. Elmalar pembedir.
2. Karıncalar yavaştır.
3. Dondurma sıcaktır.
4. Kediler fare kovalar.
5. Bebekler tüylüdür.
4. 1. Kuşlar kocamandır.
2. Motorsikletler havlar.
3. Bıçaklar yumuşaktır.
4. Bulutlar beyazdır.
5. Tavuklar yazı yazar.
5. 1. Gemiler uçar.
2. Kareler yuvarlaktır.
3. Çorabı ayağımıza giyeriz.
4. Bisikletler süt içer.
5. İnsanlar iki kulaklıdır.
6. 1. Uçaklar kanatlıdır.
2. Elmalar şarkı söyler.
3. Dağlar çok küçüktür.
4. Sandalyeler ayaklıdır.
5. Makaslar kağıt keser.

SETS OF 6

1. 1. Muzlar dişlidir.
2. Köpekler gitar çalar.
3. Bacağımız parmaklıdır.
4. Mektupları pulla göndeririz.
5. Muzlar şarıdır.
6. Kurbağalar zıplar.
2. 1. Oyuncak ayılar yumuşaktır.
2. Ördekler suda yaşar.
3. Çocuklar üç kolludur.
4. Evimiz şarkı söyler.
5. Ördekler beş ayaklıdır.
6. Kar soğuktur.
3. 1. Saatler zamanı gösterir.
2. Ayran tatlıdır.
3. Kurbağalar uzun kulaklıdır.
4. Ağaçlar müzik çalar.
5. Toplar yuvarlaktır.
6. Balıklar suda yaşar.
4. 1. Arılar sokar.
2. Koyunlar kuyrukludur.
3. İnekler uçar.
4. Köpek balığı kocamandır.
5. Bulutlar siyahtır.
6. Pamuk ağırdır.
5. 1. Ağaçlar tüylüdür.
2. Marketler yiyecek satar.
3. Domates kırmızıdır.
4. Kediler çok büyüktür.
5. Tavşanlar uzun kulaklıdır.
6. Tavuklar okula gider.
6. 1. Kirazlar mavidir.
2. Ağaçlar yapraklıdır.
3. Demir hafiftir.
4. Yılanlar zıplar.
5. Kekler tatlıdır.
6. Tekerlekler karedir.

Three versions of 'Birthday Puppy' and 'Chocolate Bar' stories with their drawings

'Birthday Puppy' story (Neutral)

Bugün Mehmet'in doğum günü ve annesi ona yavru bir köpekle sürpriz yapmak istiyor. Mehmet'in annesi yavru köpeği bodruma saklıyor. Mehmet annesine, "Anneciğim, doğum günüm için bana yavru bir köpek almanı çok istiyorum" diyor. Annesinin yavru köpekle Mehmet'e sürpriz yapmak istediğini unutma! Bu yüzden ona yavru bir köpek aldığını söylemek yerine annesi, "Üzgünüm Mehmetciğim, doğum günün için sana yavru bir köpek almadım. Onun yerine sana çok güzel bir oyuncak aldım" diyor.

Reality control question: Annesi doğum günü için Mehmet'e gerçekten ne aldı?

Şimdi Mehmet annesine "Dışarıya oynamaya çıkıyorum." diyor. Dışarıya çıkarken patenlerini almak için bodruma iniyor. Bodrumda doğum günü hediyesi yavru köpeği buluyor! Kendi kendine "Vay canına, annem bana oyuncak almamış, gerçekten doğum günüm için bana yavru bir köpek almış" diyor. Annesi Mehmet'in bodruma indiğini ve doğum günü hediyesi yavru köpeği bulduğunu görmüyor.

1st-order ignorance: Mehmet doğum günü için annesinin ona yavru bir köpek aldığını biliyor mu?

Linguistic control: Annesi Mehmet'in bodrumdaki doğum günü hediyesi yavru köpeği gördüğünü biliyor mu?

O sırada zır zır zır zır telefon çalıyor! Mehmet'in anneannesi doğum günü partisinin saat kaçta olduğunu öğrenmek için arıyor. Anneannesi telefonda Mehmet'in annesine "Mehmet doğum günü için ona gerçekten ne aldığını biliyor mu?" diye soruyor.

Şimdi hatırlayalım, Mehmet'in annesi, doğum günü için Mehmet'e aldığı şeyi Mehmet'in gördüğünü bilmiyor. Daha sonra anneanne Mehmet'in annesine "Mehmet doğum günü için ona ne aldığını düşünüyor?" diye soruyor.

2nd- order false belief: Mehmet'in annesi anneanneye ne cevap verir?

Justification: Mehmet'in annesi neden böyle bir cevap verir?

'Birthday Puppy' story (-DI)

Dün Mehmetlerdeydim. Mehmet'in doğum günüydü ve annesi ona yavru bir köpekle sürpriz yapmak istedi. Mehmet'in annesi yavru köpeği bodruma sakladı. Mehmet annesine, "Anneciğim, doğum günüm için bana yavru bir köpek almanı çok istiyorum" dedi. Annesinin yavru köpekle Mehmet'e sürpriz yapmak istediğini unutma! Bu yüzden ona yavru bir köpek aldığını söylemek yerine annesi, "Üzgünüm Mehmetciğim, doğum günün için sana yavru bir köpek almadım. Onun yerine sana çok güzel bir oyuncak aldım" dedi.

Reality control question: Annesi doğum günü için Mehmet'e gerçekten ne aldı?

Mehmet annesine "Dışarıya oynamaya çıkıyorum." dedi. Dışarıya çıkarken patenlerini almak için bodruma indi. Bodrumda doğum günü hediyesi yavru köpeği buldu! Kendi kendine "Vay canına, annem bana oyuncak almamış, gerçekten doğum günüm için bana yavru bir köpek almış" dedi. Annesi Mehmet'in bodruma indiğini ve doğum günü hediyesi yavru köpeği bulduğunu görmedi.

1st-order ignorance: Mehmet doğum günü için annesinin ona yavru bir köpek aldığını biliyor muydu?

Linguistic control: Annesi Mehmet'in bodrumdaki doğum günü hediyesi yavru köpeği gördüğünü biliyor muydu?

O sırada zır zır zır zır telefon çaldı! Mehmet'in anneannesi doğum günü partisinin saat kaçta olduğunu öğrenmek için aradı. Anneannesi telefonda Mehmet'in annesine "Mehmet doğum günü için ona gerçekten ne aldığını biliyor mu?" diye sordu.

Şimdi hatırlayalım, Mehmet'in annesi, doğum günü için Mehmet'e aldığı şeyi Mehmet'in gördüğünü bilmiyordu. Daha sonra anneanne Mehmet'in annesine "Mehmet doğum günü için ona ne aldığını düşünüyor?" diye sordu.

2nd-order false belief: Mehmet'in annesi anneanneye ne cevap verdi?

Justification: Mehmet'in annesi neden böyle bir cevap verdi?

'Birthday Puppy' story (-MIŞ)

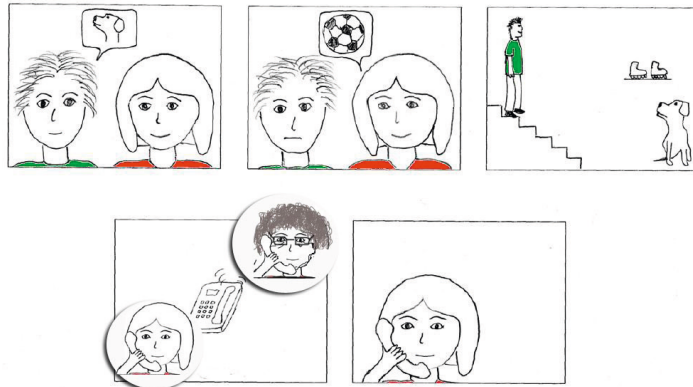
Bak Mehmet. Geçen hafta Mehmet'in doğum günüymüş. Annesi ona yavru bir köpek sürpriz yapmak istemiş. Mehmet'in annesi yavru köpeği bodruma saklamış. Mehmet annesine, "Anneciğim, doğum günüm için bana yavru bir köpek almanı çok istiyorum" demiş. Annesinin yavru köpek Mehmet'e sürpriz yapmak istediğini unutma! Bu yüzden ona yavru bir köpek aldığını söylemek yerine annesi, "Üzgünüm Mehmetciğim, doğum günün için sana yavru bir köpek almadım. Onun yerine sana çok güzel bir oyuncak aldım" demiş.

Reality control question: Annesi doğum günü için Mehmet'e gerçekten ne almış?

Mehmet annesine "Dışarıya oynamaya çıkıyorum." demiş. Dışarıya çıkarken patenlerini almak için bodruma inmiş. Bodrumda doğum günü hediyesi yavru köpeği bulmuş! Kendi kendine "Vay canına, annem bana oyuncak almamış, gerçekten doğum günüm için bana yavru bir köpek almış" demiş. Annesi Mehmet'in bodruma indiğini ve doğum günü hediyesi yavru köpeği bulduğunu görmemiş.

1st-order ignorance: Mehmet doğum günü için annesinin ona yavru bir köpek aldığını biliyor muymuş?

Linguistic control: Annesi Mehmet'in bodrumdaki doğum günü hediyesi yavru köpeği gördüğünü biliyor muymuş?



Adapted from Flobbe et al. (2009) under a CC BY license, with permission from the authors. The grandmother image was added to the original drawing in order to make the story more explicit.

O sırada zır zır zır zır telefon çalmış! Mehmet'in anneannesi doğum günü partisinin saat kaçta olduğunu öğrenmek için aramış. Anneannesi telefonda Mehmet'in annesine "Mehmet doğum günü için ona gerçekten ne aldığını biliyor mu?" diye sormuş.

Şimdi hatırlayalım, Mehmet'in annesi, doğum günü için Mehmet'e aldığı şeyi Mehmet'in gördüğünü bilmiyormuş. Daha sonra anneanne Mehmet'in annesine "Mehmet doğum günü için ona ne aldığını düşünüyor?" diye sormuş.

2nd-order false belief: Mehmet'in annesi anneanneye ne cevap vermiş?

Justification: Mehmet'in annesi neden böyle bir cevap vermiş?

'Chocolate Bar' story (NEUTRAL)

Bak, bunlar Can ile Ece kardeşler. Oturma odasında oynuyorlar. Biraz sonra anneleri alışverişten dönüyor, torbadan bir paket çikolata çıkarıyor. Çikolatayı Can'a veriyor. Ece'ye hiç çikolata vermiyor çünkü yaramazlık yapıyor. Can çikolatanın birazını yiyor ve kalanını çekmeceye koyuyor. Ece'ye hiç çikolata vermiyor. Ece da buna çok sinirleniyor. Can mutfığa annesine yardım etmek için bulaşıkları yıkamaya gidiyor. Ece oturma odasında tek başına oturuyor. Can ise mutfakta. Ece Can'a sinirlendiği için çikolatayı saklıyor. Çikolatayı çekmecedan alıyor ve oyuncak sandığına koyuyor. Can bulaşıkları yıkamakla meşgul. Can, meyve kabuklarını bahçedeki çöp kovasına atmaya giderken pencereden oturma odasını görüyor. Ece'nin çikolatayı çekmecedan alıp oyuncak sandığına koyduğunu görüyor. Ece ise Can'ı görmüyor.

Reality control question: Çikolata şimdi nerede?

1st-order ignorance: Can, Ece'nin çikolatayı oyuncak sandığına sakladığını biliyor mu?

Linguistic control: Ece çikolatayı saklarken Can'ın onu gördüğünü biliyor mu?

Can bulaşıkları bitiriyor. Karnı acıkıyor. Çikolatasından biraz yemek istiyor. Can oturma odasına giriyor. "Canım biraz çikolata istiyor." diyor.

2nd-order false belief: Ece çikolata için Can'ın nereye bakacağını düşünüyor?

Justification: Ece neden böyle düşünüyor?

'Chocolate Bar' story (-DI)

Bak, bunlar Can ile Ece kardeşler. Geçen gün onların evindeydim. Oturma odasında oynuyorlardı. Biraz sonra anneleri alışverişten döndü, torbadan bir paket çikolata çıkardı. Çikolatayı Can'a verdi. Ece'ye hiç çikolata vermedi çünkü yaramazlık yapıyordu. Can çikolatanın birazını yedi ve kalanını çekmeceye koydu. Ece'ye hiç çikolata vermedi. Ece da buna çok sinirlendi. Can mutfığa anesine yardım etmek için bulaşıkları yıkamaya gitti. Ece oturma odasında tek başına oturuyordu. Can ise mutfaktaydı. Ece Can'a sinirlendiği için çikolatayı sakladı. Çikolatayı çekmecedan aldı ve oyuncak sandığına koydu. Can bulaşıkları yıkamakla meşguldü. Can, meyve kabuklarını bahçedeki çöp kovasına atmaya giderken penceren oturma odasını görüyordu. Ece'nin çikolatayı çekmecedan alıp oyuncak sandığına koyduğunu gördü. Ece ise Can'ı görmedi.

Reality control question: Çikolata neredeydi?

1st-order ignorance: Can, Ece'nin çikolatayı oyuncak sandığına sakladığını biliyor muydu?

Linguistic control: Ece çikolatayı saklarken Can'ın onu gördüğünü biliyor muydu?

Can bulaşıkları bitirdi. Karnı acıktı. Çikolatasından biraz yemek istedi. Can oturma odasına gitti. "Canım biraz çikolata istiyor." dedi.

2nd-order false belief: Ece çikolata için Can'ın nereye bakacağını düşündü?

Justification: Ece neden böyle düşündü?

'Chocolate Bar' story (-MIŞ)

Bak, bunlar Can ile Ece kardeşler. Geçenlerde Can ile Ece oturma odasında oynuyorlarmış. Biraz sonra anneleri alışverişten dönmüş, torbadan bir paket çikolata çıkarmış. Çikolatayı Can'a vermiş. Ece'ye hiç çikolata vermemiş çünkü yaramazlık yapıyorlarmış. Can çikolatanın birazını yemiş ve kalanını çekmeceye koymuş. Ece'ye hiç çikolata vermemiş. Ece da buna çok sinirlenmiş. Can mutfığa anesine yardım etmek için bulaşıkları yıkamaya gitmiş. Ece oturma odasında tek başına oturuyormuş. Can ise mutfaktaymış. Ece Can'a sinirlendiği için çikolatayı saklamış. Çikolatayı çekmecedan almış ve oyuncak sandığına koymuş. Can bulaşıkları yıkamakla meşgulmüş. Can, meyve kabuklarını bahçedeki çöp kovasına

atmaya giderken pencereden oturma odasını görüyormuş. Ece'nin çikolatayı çekmecedan alıp oyuncak sandığına koyduğunu görmüş. Ece ise Can'ı görmemiş.

Reality control question: Çikolata neredeymiş?

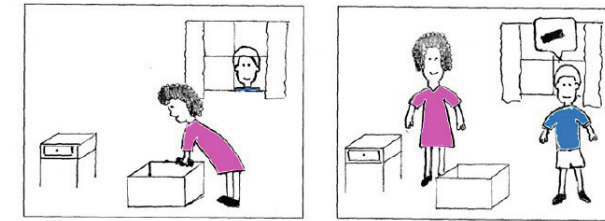
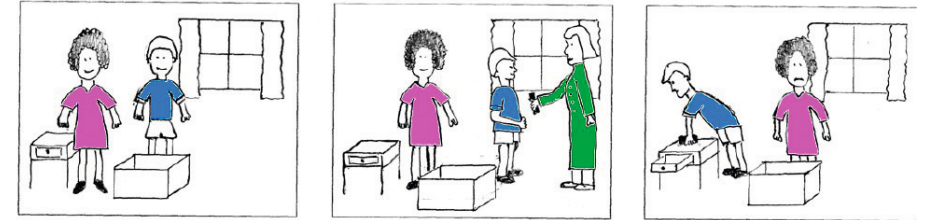
1st-order ignorance: Can, Ece'nin çikolatayı oyuncak sandığına sakladığını biliyor muymuş?

Linguistic control: Ece çikolatayı saklarken Can'ın onu gördüğünü biliyor muymuş?

Can bulaşıkları bitirmiş. Karnı acıkmış. Çikolatasından biraz yemek istemiş. Can oturma odasına gitmiş. "Canım biraz çikolata istiyor." demiş.

2nd-order false belief: Ece çikolata için Can'ın nereye bakacağını düşünmüş?

Justification: Ece neden böyle düşünmüş?



Adapted from Flobbe et al. (2009) under a CC BY license, with permission from the authors.

Publication List

Journal Articles

- Arslan, B.**, Taatgen, N., & Verbrugge, R. (2017). Five-year-olds' systematic errors in second-order false belief tasks are due to first-order theory of mind strategy selection: A computational modeling study. *Frontiers in Psychology*, 8:275. <https://doi.org/10.3389/fpsyg.2017.00275>
- Arslan, B.**, Hohenberger, A., & Verbrugge, R. (2017). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PLOS ONE* 12(1): e0169510. doi: 10.1371/journal.pone.0169510
- Arslan, B.**, Verbrugge, R., & Taatgen, N. (2017). Cognitive control explains the mutual transfer between dimensional change card sorting and first-order false belief understanding: A computational modeling study on transfer of skills. *Biologically Inspired Cognitive Architectures*. <http://dx.doi.org/10.1016/j.bica.2017.03.001>
- Arslan, B.**, Verbrugge, R., Taatgen, N., Hollebrandse, B. (in preparation). Five-year-old children's development of second-order false belief reasoning can be accelerated: An evaluation of different feedback methods.

Peer-reviewed Conference Papers

- Arslan, B.**, Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2015). Teaching children to attribute second-order false beliefs: A training study with feedback. In D. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 108-113). Austin, TX: Cognitive Science Society.
- Arslan, B.**, Wierda, S., Taatgen, N., & Verbrugge, R. (2015). The role of simple and complex working memory strategies in the development of first-order false belief reasoning: A computational model of transfer of skills. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.), *Proceedings of the 13th International Conference on Cognitive Modeling* (pp. 100-105). Groningen: University of Groningen.
- Arslan, B.**, Taatgen, N., & Verbrugge, R. (2013). Modeling developmental transitions in reasoning about false beliefs of others. In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling* (pp. 77-82). Ottawa: Carleton University.

- Arslan, B.**, Hohenberger, A., & Verbrugge, R. (2012). The development of second-order social cognition and its relation with complex language understanding and working memory. In N. Miyake (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp., 1290-1295). Sapporo: Japan.
- Gokcay, D., Cakir, M., Gozenman, F., **Arslan, B.**, Ciftcioglu, M., Taskin, & K., Ayaz, H. (2011). Changing pattern search component during probability learning: a study using fNIR neuroimaging. *17th Annual Meeting of the Human Brain Mapping* (pp., 26-30). Quebec: Canada.

Extended Abstracts

- Arslan, B.**, Verbrugge, R., Taatgen, N. & Hollebrandse, B. (2014). Teaching children to attribute second-order false belief: A training study. In J. Szymanik & R. Verbrugge (Eds.), *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives* (Vol. 1208, pp. 1-5). Groningen: CEUR Workshops Proceedings.
- Wierda, S. & **Arslan, B.** (2014). Modeling theory of mind in ACTransfer. In J. Szymanik & R. Verbrugge (Eds.), *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives* (pp., 40-44). Groningen: CEUR Workshops Proceedings.

Acknowledgments

Dedicating four years of your life to a specific goal has both psychological and technical challenges. I was extremely lucky that I was surrounded with great supervisors, colleagues, friends and family who have always been supportive. Otherwise, neither this dissertation could have been in its current form nor could my mental state. I hope I have found a way to show my appreciation over the years but hereby I would like to take the chance to thank them all officially.

First of all, I would like to thank my supervisors Rineke Verbrugge and Niels Taatgen for their continuous support:

Rineke, it would take me to write another book if I tried to acknowledge all of your contributions to this dissertation and to my scientific and personal growth. I am sure that all the people who know you and reading this acknowledgement would agree with me that you are a role model for many people in academia both in terms of your scientific achievements and also in terms of your personal qualities. It was a privilege to be your PhD student. I have learned a lot from you. You always made sure that I improved my technical and transferable skills. I started my PhD with a great motivation; however, there were dark moments that I was discouraged by the academic review system but your positive attitude and support made me perseverant. You always found time to read and give feedback to the drafts of our papers repeatedly. In addition to the explicit feedback you gave, I took every chance to learn from you by observing you in our meetings and in other events. While you were involved at each step of this dissertation, you gave me the freedom to develop and implement my own ideas. You supported me to attend many national and international conferences and workshops, which helped me expand my network. I do not know how you initially decided to offer me a PhD position at your group; however, I do know that carrying out research under your supervision is the best thing that happened to me and I tried my best not to disappoint you. Thank you for “everything” [that you have done for me].

Niels, the first course that I took in my MSc in Cognitive Science programme was “Computational Models of Mind” and the first couple of papers that I read were yours and John Anderson’s. Those papers were what made me fall in love with computational cognitive modeling. I was amazed by your perspective on complex things and how simply you explained them. Later on, I had the chance to be your PhD student, thanks to Rineke. From the first day that I started my PhD, I was hung up on your every word during our meetings and CogMod meetings in

order to learn from you as much as possible. Although we haven't spent so many hours working together as much as I spent with Rineke, your input was always very useful and to the point. You were and still are one of my scientific idols.

There is someone that I would like to thank right after my supervisors: Harmen de Weerd (a.k.a. Harmish the General Problem Solver). My unofficial daily supervisor, paranymp, colleague, friend, and Zumba buddy. During these four years, I do not know how many times I came to your office to ask technical questions, to get your opinion about my research, to discuss my results, papers, or just to share something that I was happy or sad about, to kindly ask you to translate the stories for my experiment and the *Samenvatting* of this dissertation from English to Dutch and vice versa. There was not even one instance out of those uncountable instances that you did not have time for me. More importantly, you never made me regret that I got your opinion and I shared something with you. You are a very smart and intellectual person. However, these features are nothing compared to how kind, humble, gentleman and open minded you are. I hope life will treat you well and you will always get what you want. You should know that you have unlimited credit on my mental card and I will be there whenever you need.

There is also another person who has supported me since my master's thesis and who deserves special acknowledgement – Annette Hohenberger. Annette, we have been working closely for six years by now and your support was invaluable. I have started to learn about cognitive development and human memory from your courses. You supervised my master's thesis together with Rineke and we continued to work together during the last four years in order to publish a paper in which you were one of the co-authors. You, Rineke and I had many meetings together over Skype. During those meetings, you always encouraged me and always provided useful comments. We got upset together when we received rejections and we celebrated together the good moments. You always found time to write reference letters for me and I was moved when I read them. I am very lucky to know you. Thank you!

I also would like to thank my reading committee members Petra Hendriks, Maartje Raaijmakers, and Lael Schooler for spending their valuable time to read my dissertation.

I was very lucky to be a member of two groups in our department: Social Cognition Research group and Cognitive Modeling group. I want to thank the members of our Social Cognition Research group Ben Meijering, Daniël van der Post, Jakub Szymanik, Harmen de Weerd, Stefan Wierda, and Rineke Verbrugge. During our

meetings, we read and criticized many papers about theory of mind from different perspectives, such as logic, comparative psychology, multi-agent systems, developmental psychology and computational agents. I have learned a lot from all of your comments and insights. I also want to thank Cognitive Modeling group members Ben Meijering, Christopher Stevens, Enkhbold Nyamsuren, Fokie Cnossen, Hagit Shaposhnik, Hedderik van Rijn, Ioanna Katidioti, Jacolien van Rij, Jelmer Borst, Katja Mehlhorn, Katja Paul, Margreet Vogelzang, Marieke van Vugt, Marijke Beulen, Menno Nijboer, Niels Taatgen, Stefan Huijser, Stefan Wierda, Oscar Valmuel and Trudy Buwalda for their useful comments about my research as well as my presentation skills. I would especially like to thank Trudy, Harmen, Jelmer, Niels, and Rineke for joining my mock interviews and providing me useful feedback repeatedly and thanks to Enkhbold and Ben for their useful comments on my first computational cognitive model when I started my PhD.

During the last four-years-and-a-half, I had different officemates. I want to thank Jakub, Bea, Ben, Sujata, Gyuhee, and Yuri for being very nice company in the office.

One of our experiments took almost two years starting from designing the experiment to the completion of testing the children. It would be impossible to realize our training study without the support of Bart Hollebrandse, who is one of the co-authors of our training paper and who helped us to find a school to run our experiments, and kindly accepted our request to use his version of second-order false belief task. Thanks to the Joseph Haydn School in Groningen and the students' families who allowed us to conduct our training study; Avik Kumar Maitra, who is the illustrator of the stories; Maximilian Seidler, Amélie la Roi, Bram Wiggers, Marlies Hoekstra, Marten Schutten, Merel Wiersma, who tested children; and last but not least Bea van Meerveld, who is the voice of the stories that was used in our training study. We spent many hours together in the studio, Bea. We also shared an office for some time and had dinners together. It was very nice to have you around.

I also would like to thank John Anderson and Christian Lebiere for inviting me to Carnegie Mellon University and for their hospitality during my short visit to Pittsburgh, and to thank Educational Testing Service R&D team for their hospitality during my short visit to Princeton. Thanks to Irina Polyanskaya, Torben Braüner and Patrick Blackburn for sharing their comments about our work, for inviting me to their workshop in Denmark and for their hospitality.

I would like to thank Jennifer Spenader, Nicole Baars-Verbrugge and Stefani Nellen for inviting me to their houses in multiple occasions and making me feel

very welcome. I especially would like to thank Jennifer for giving me a hug in one of my darkest moments and for knocking on my door from time to time just to say good morning or to share something that made her happy.

I want to thank my former and current colleagues/friends in Artificial Intelligence Department Ana Bosnic, Aswin van Woudenberg, Ben Wolf, Bowornrat Sriman, Faik Karaaba, Marco Weiring, Sheng He, Stippe Pandzic, Olarik Surinta, and Pry Mania for all the lunches and coffees that we had together. Special thanks go to Ana for all the other fun times we had together.

I would like to thank faculty secretaries Elina Sietsema and Carlijne de Vries, who were always very helpful and kind. I really appreciate all your support.

Joanna Smolonska who worked for the cover and layout design of this dissertation also deserves special thanks. Thank you for taking care of all the little details patiently.

Being an activist has never been one of my characteristics. That is why, I always appreciate the people who speak up for improvements. I would like to take the chance of thanking Catarina Dutilh Novaes, Deniz Başkent, Fokie Cnossen, Justine Jacot, Nina Gierasimczuk, Olivia Guest, Petra Hendriks, and Rineke Verbrugge who put effort to make women visible in academia. I especially would like to thank Olivia Guest. Even though we met only once five years ago at a conference, your friendship and support has been significant.

I also would like to thank my former professors from my master's studies at the Middle East Technical University; Didem Gökçay, Murat Perit Çakır, Ceyhan Temürcü, Bilge Say, Cem Bozşahin, and Annette Hohenberger from whom I have learned a lot about cognitive science.

After spending many hours at work by thinking about new ideas and trying to improve your technical skills, it is very important to have a relaxing environment at home and not to have any troubles. I would like to thank my landlord, John van Bergen, for his contributions to keep my mind in piece during my three and a half years at his place. Probably, it is the first time that a PhD thesis acknowledgment includes a landlord and you might think that I overdid it. However, considering the fact that another landlord destroyed my mental health, which had an impact on my academic performance, I think it is important to thank John for being a great landlord.

Bearing in mind that I have spent more than 30 years of my life in a big city, living in the city centre of Groningen was an amazing experience for many reasons. No stress because of the traffic jam; after stepping out from your apartment all the shops/cafes/bars are around you; you always come across a friend and have unplanned drinks together; you get your coffee without saying a word except "hi", thanks to Starbucks barista; you have a specific store that knows your fashion style and you get what you want without even looking at the options, thanks to the ladies at the Betties Creative Casuals.

As an only child, I have always been into my friends more than an average person. Living in Ankara for 32 years, I have been friends with Anıl, Can, Nihan, Serhan, Sertaç (*karrrrdeşşşimmm*) and Ülge more than 10 years and with most of them more than 20 years. We had spent almost all of our free time together, we had been together when we had our first experiences in life, we had died from laughing together and then came back to life from our ashes. You have all significantly contributed to the person who I have become. Although we are living far away and we don't see each other often, knowing that you are and will always be there for me makes me feel at home wherever I go. Special thanks go to Ülge.

Ülge, it could have not been possible to be titled as a PhD or even an MSc without your support. You have always believed in me more than I have believed in myself and you have always supported me (or forced me ☺) to push the boundaries of my knowledge and skills. Starting from my master's studies, you have corrected uncountable emails, assignments, reports and papers that I wrote, both in terms of their content and also in terms of language. If someone asks you what theory of mind is, I know that you will give a definition by heart. Not only intellectually but in many different aspects of my life you always make sure that I have the best possible outcome including but not limited to the effective usage of my 20m2 place. You visited me here multiple times and you have become a real fan of Groningen, *Martinitoren*, *oliebollen* and *bitterballen*. I don't know how you managed but the most of the shop owners recognize you and give you free products, such as *oliebollen*. We travelled to many countries in Europe and to the United States together and accidentally hit our heads to each other's multiple times (dumb and dumber ☺). We had talked over the phone 10¹⁰⁰ hours about politics, philosophy, art, design, caricatures – to which I have never been able to laugh except a couple of those, artificial intelligence, economics, society, geography, history etc. If we had published online all those conversations as a podcast, I think, we might have had many followers. In addition to the intellectual and fun times, you have always listened to all of my problems, tried to calm me down when I was sad or angry, and sacrificed yourself and became my stress ball time to time.

Importantly, you also visited, called and helped my family when I was away from them. You are the most intellectual, diligent, gracious and brave (☺)person that I have ever known. Thank you for putting up with me. Prepare yourself for the long trips to the US. *Cap toujours!*)(*

Leaving all the above-mentioned friends behind and starting a new life in Groningen was an emotional challenge. However, there is one person, who has such a unique personality, made this challenge significantly easier starting from the day 1 – Ioanna Katidioti (a.k.a Greek Goddess). When I came to our department for the first time, you were the one who was really interested in who I am. You were the one who first invited me for dinners and drinks. You were the one who immediately understood what I wanted to say even though I wasn't used to speaking in English fluently. You introduced me to your boyfriend (*Morak*, so cu-uute!), friends, family and Zizi, who have also become my *Morak*, my friends, my family and my Zizi over time. You made me laugh 24/7 (or 7/24, or 74/7) at work by your presentation skills and at other times by your unexpected comments in between conversations (e.g., London, PhD, *vloxoutsu*). You were the one who bought me dolma without any reason (except you love me a lot!) and brought me presents whenever you went abroad. You were the one who gave me some extra points and took some of them back time to time. You were the manager of the line. You were the one who gave me limited number of coupons for literally scratching my back but who scratched my back unlimitedly anyway. We blurred together our vision and said “*volgende volgende*” when it was necessary. We were each other's +1s in many social activities. We went to spa and massages together and I impressed you with my dolphin skills. We entertained each other at Zumba, so on and so forth. In summary, you made both my work and my social life more fun. The fun times that we spent together aside, you have always been there when I was upset. I still do not know how you manage to be with me whenever I need when you are reading more than 60 books a year, going to Zumba, Capoeira and dance classes five times a week, watching all the TV series, and joining almost all social events. Although you don't literally put it into words, you always found a way to make me feel special. If I would write a novel about our times together until now, the title would be as follows: ‘Greek Goddess facilitates and I predict the development of Groningen: An evaluation of fashion weeks.’ To conclude, I feel very lucky to have you in my life and I really appreciate that you appreciate that I appreciate you. As you predicted in your acknowledgements, it seems like we will be indeed in the other parts of the world – at least for a while. However, Shante, you stay [with me].

I also would like to thank to Stephan (a.k.a. *Morak*) for visiting me when I was sick, for coming to my place to catch a mouse immediately after I called him, for our many drinks, and dinners together.

Now, it is time to thank my other Greeks: Fotini, Konstantina, and Sissy. We had a lot of fun together and shared so many things. I always enjoyed spending time with you. You definitely made my time in Groningen a lot better. I am glad to know you all!

Last but not least, I would like to thank Stephanie for our traditional Friday evening RuPaul sessions, for being there for me at 3 a.m. when I first saw a mouse at my place and for letting me crash at her place, for our nice trip together to Lisbon, for many dinners, drinks and *borrels* together and for sharing her interesting stories and making me laugh. I am glad to know you!

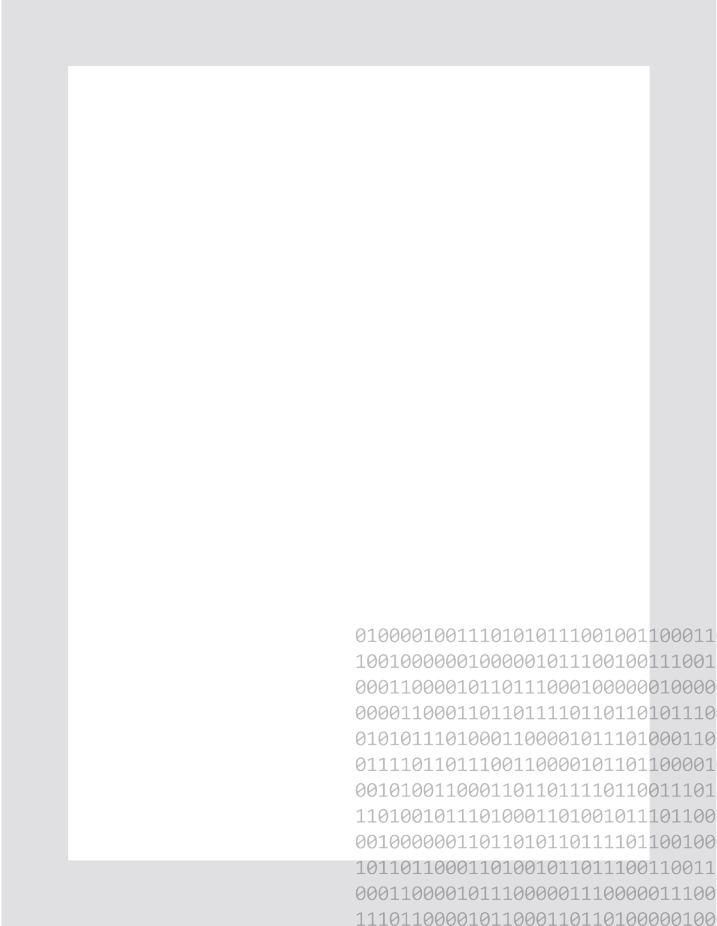
I would like to thank Ahmet & Cansu, Alex, Aliki & Yiannis, Andre, Arman, Christos, Cyrus, Eva, Guilia, Gyhuue & Michiel, Ina, Juan, Meniz & Yiğit, Michela, Pınar & Seçkin, Reto & Tonia, and Turhan for all the drinks, dinner, and parties we had together.

I also want to thank to Konstantina, Neşe, Setenay, Sertaç, Tunahan, Yasemin & Deniz for visiting me in Groningen.

I would like to finalize this long Acknowledgments section by thanking to my mother Güzin, my father Veysel, my aunt Deniz, my uncle Metay, my cousin Tunahan and my grandparents who have always supported me and loved me unconditionally. I especially thank my parents for all their sacrifices in order to make sure that I had the best education that I could get and I had all the other opportunities that they couldn't have; for never judging who I am; for letting me be out of the norms and for their trust in me. I dedicated this dissertation to my grandparents, Sevinç & Şeyh Ömer Kal and Nazlı & Mehmet Arslan. Unfortunately, only one of my grandparents is alive to read this and be proud of me. Grandmother Sevinç, you are the best and most fun grandmother for sure! I love you all!

Bu uzun Teşekkür bölümünü beni daima desteklemiş ve koşulsuz olarak sevmiş olan annem, babam, teyzem, eniştem ve kuzenim ile anneannem, babaannem ve dedelerime teşekkür ederek tamamlamak istiyorum. Anneme ve babama, mümkün olan en iyi eğitimi almamı ve kendilerinin sahip olamadığı diğer tüm olanaklara sahip olmamı sağlamak adına yapmış oldukları tüm fedakârlıklar için; beni asla yargılamadıkları

için; alışılmışın dışında olmama izin verdikleri için ve bana olan güvenleri için özel olarak teşekkür ediyorum. Doktora tezimi anneannem Sevinç Kal ve dedem Şeyh Ömer Kal ile babaannem Nazlı Arslan ve dedem Mehmet Arslan'a adıyorum. Ne yazık ki içlerinden sadece anneannem bu satırları okuyup gurur duyabilecek. Anneanneeeeeee, sen bir insanın sahip olabileceği en mükemmel ve en eğlenceli anneannesin! Hepinizi çok seviyorum!



010000100111010101110010011000110111010
100100000010000010111001001110011011011
000110000101101110001000000100000100100
000011000110110111101101101011100000111
010101110100011000010111010001101001011
011110110111001100001011011000010000000
001010011000110110111101100111011011100
110100101110100011010010111011001100101
001000000110110101101111011001000110010
101101100011010010110111001100111001000
000110000101110000011100000111001001101
111011000010110001101101000001000000000
101001110100011011110010000001110100011
010000110010100100000011001000110010101
110110011001010110110001101111011100000
110110101100101011011100111010000100000
011011110110011000100000000010100111001
101100101011000110110111101101110011001
000010110101101111011100100110010001100
101011100100010000000001010011101000110
100001100101011011110111001001111001001
000000110111101100110001000000110110101
1010010110111001100100